# ESTADÍSTICA DESCRIPTIVA Y PROBABILIDAD PARA LAS CIENCIAS DE LA INFORMACIÓN CON EL USO DEL SPSS

FAUSTO FRANCISCO MATOS URIBE
FORTUNATO CONTRERAS CONTRERAS
JULIO CESAR OLAYA GUERRERO

ASOCIACIÓN DE BIBLIOTECÓLOGOS DEL PERÚ

Editado por la:

ASOCIACIÓN DE BIBLIOTECÓLOGOS DEL PERÚ
Calle Hipólito Bernardette Nº 106 - Barranco
Lima - Perú

Primera edición electrónica, setiembre 2020.

Libro electrónico disponible en:

https://archive.org/

ISBN: 978-612-48342-0-2

**AGRADECIMIENTOS**

A NUESTROS PADRES Y HERMANOS
A LOS ESTUDIANTES Y COLEGAS



# Contenido

# PRÓLOGO

El presente libro es el producto de los apuntes de clases dictado en diversas instituciones de educación superior a lo largo de varios años, teniendo como pre-requisitos el curso de matemática básica.

El objetivo de esta obra es presentar métodos y procedimientos para el análisis de datos cuando estos por su naturaleza presentan incertidumbre en la toma de decisión.

El libro consta de 10 capítulos, acompañados de ejercicios resueltos y propuestos con la aplicación del SPSS.

Esperamos que el presente texto de consulta sirva de guía al estudiante o investigador, para comprender los métodos que usa esta disciplina para analizar datos.

**LOS AUTORES**

# CAPÍTULO 1

## INTRODUCCIÓN

Este primer CAPÍTULO tiene por objeto presentar algunos conceptos básicos de estadística, asi como de sumatorias simples y dobles los cuales se van a utilizar para una mejor comprensión de los capítulos siguientes.

## CONCEPTOS BÁSICOS DE ESTADÍSTICA Y SUMATORIAS

### 1.1. ESTADÍSTICA

Es una ciencia que proporciona métodos y procedimientos para el análisis de datos, cuando éstos por su naturaleza presenta incertidumbre en la toma de decisión.

Otras definiciones:

Es una matemática aplicada para el análisis de datos.

Es el estudio de los fenómenos aleatorios.

Es la ciencia que se encarga de la recolección, ordenamiento, representación, análisis e interpretación de datos generados en una investigación sobre hechos, individuos o grupos de los mismos, para deducir de ello conclusiones precisas o estimaciones futuras (Salazar, Castillo, 2018, p.13).

### 1.2. RAMAS DE LA ESTADÍSTICA

La estadística se divide en dos ramas:

- ESTADÍSTICA DESCRIPTIVA

Es parte de la estadística que consiste en la recolección, organización, presentación, análisis e interpretación de un conjunto de datos para una o más variables de interés del investigador.

Es la ciencia mediante la cual se recopila, organiza, presenta, analiza e interpreta datos de manera informativa tal que describa fácil y rápidamente las características esenciales de dichos datos mediante el empleo de diversos métodos gráficos, tabulares o numéricos, etc. (Suarez, 2018, p.13).

- ESTADÍSTICA INFERENCIAL

Es parte de la estadística que se encarga en tomar decisiones sobre el comportamiento de una variable de interés definida en la población en estudio; utilizando para tal objetivo una muestra al azar. Es decir a partir de los resultados obtenidos al procesar los datos de una muestra, se generaliza los resultados de dicha variable en la población, con un cierto riesgo que es medido por la teoría de la probabilidad. La estadística inferencial proporciona métodos y procedimientos para el análisis de datos.

Llamada también inferencia estadística, la cual consiste en llegar a obtener conclusiones o generalizaciones de la población a partir de una muestra de ella, es decir, emplea métodos para buscar información que permita determinar propiedades de la población basándose en el estudio de los datos de una muestra tomada a partir de ella (Suarez, 2018, p.13),

## 1.3. CONCEPTOS BÁSICOS

### 1.3.1. Población

Es la totalidad de unidades elementales (personas, animales u objetos) sujeto a estudio y que poseen una característica común; al tamaño de la población (número de unidades elementales) se simboliza con la letra mayúscula “N”. Una población en estudio puede ser finita (se pueden enumerar sus elementos) o infinita (no se pueden enumerar sus elementos).

Ejemplos:

a. Todas las tiendas que se dedican a la venta de gas en el distrito de la Molina (N = 650)
b. Todos los establos de ganado vacuno en el departamento de Cajamarca (N =1,350)

c. Todos los Alumnos matriculados en el semestre 2000-II, en la UNMSM.
d. Todos los Cadetes de la Escuela de Oficiales de la PNP, en proceso de formación.
e. Todas las papeletas impuestas por la PNP, por infringir al Reglamento Nacional de Tránsito, durante el mes de diciembre de un año determinado (N = 2,520).
f. Todos los usuarios que concurren a la Biblioteca de Letras y Ciencias Humanas de la UNMSM, un día cualquiera. (N = 1,742)
g. Todas las estrellas del universo.

1.3.2. Unidad Elemental

Es cada uno de los elementos que conforman la población en estudios, de la cual se requiere información. En una unidad elemental se definen muchas variables de interés para el investigador.

Ejemplo

De los ejemplos anteriores de poblaciones se tienen las siguientes unidades elementales:

a. Una tienda que se dedican a la venta de gas en el distrito de la Molina.
b. Un establo de ganado vacuno en el departamento de Cajamarca.
c. Un alumno matriculado en el semestre 2000-II, en la UNMSM.
d. Un Cadete de la Escuela de Oficiales de la PNP, en proceso de formación en el año 2004.
e. Una papeleta impuesta por infringir al Reglamento Nacional de Tránsito impuestas por la Policía en el mes de Setiembre.
f. Un usuario que concurre a la Biblioteca de Letras y Ciencias Humanas de la UNMSM, un día cualquiera.
g. Una estrella del universo.

1.3.3. Dato u observación

Es el valor (atributo, conteo o medición) que toma la variable de interés en la unidad elemental.

Los valores que toma una variable en particular $X$ , se representan con letras minúsculas acompañadas por subíndice; es decir,

$$X = \{x_1, x_2, x_3, \ldots x_n\}$$

$x_i$ = valor que toma la variable $X$ en la i-ésima observación

1.3.4. Variable

Es una característica de interés que se desea estudiar en la unidad elemental. La variable en estudios la fija el investigador; el valor que toma la variable (dato u observación) cambia de unidad elemental a otra. A una variable se le simboliza con las últimas letras del abecedario, tales como X, Y, Z, o también letras mayúsculas acompañada con subíndice, tales como: $X_1, X_2, \ldots, X_k$ .

Ejemplos de variables en estudio:

1.7. X = Montos(S/.) por venta diaria de contenidos de balones de gas.
1.8. Y = Numero de ganado vacuno por establos
1.9. Z = Color de los ojos de los Cadetes de la EO-PNP
1.10. W= Volumen ($cm^3$) de contenido de liquido de una botella.
1.11. T =Tiempo (minutos) que permanece un usuario en un cajero automático.

1.3.5. Tipos de variables:

De acuerdo a los valores que toma la variable en estudio, se clasifican en:

- Variable Cualitativa

Cuando los valores que toma la variable en estudio provienen de una cualidad o atributo, se divide a su vez en nominal y jerárquica (ordinal).

- <u>Variable Cualitativa Nominal.</u>

Una variable se considera cualitativa nominal, cuando con los valores que toma esta variable no es posible establecer un orden de acuerdo a su importancia.

<u>Ejemplos</u>:

a. X = Color de los ojos de los alumnos de la Facultad de Letras y Ciencias Humanas de la Universidad Nacional Mayor de San Marcos (marrones, pardos, verdes, negros,…)
b. Y = Preferencias a determinadas marcas de jabones de tocador por las alumnas de la facultad de Letras y Ciencias Humanas de la Universidad Nacional Mayor de San Marcos.
c. Z = Sexo de los Cadetes de la EO-PNP.

- <u>Variable Cualitativa Jerárquica u Ordinal</u>

Una variable se considera cualitativa jerárquica u ordinal, cuando con los valores que toma la variable en estudio es posible establecer un orden de acuerdo a su importancia (ascendentemente o descendentemente)

<u>Ejemplos</u>:
a. X = Grados Académicos de los docentes de la Facultad de Ingeniería de Sistema de la Universidad Nacional Mayor de San Marcos: bachiller, magíster, doctor, Ph.D.
b. Y = Grados del personal de Oficiales de la Policía Nacional del Perú: Teniente General, General, Coronel, Comandante, Mayor, Capitán, Teniente, Alférez.
c. Z = Clase social: alta, media, baja

- <u>Variable cuantitativa</u>

Los valores que toma la variable provienen de un conteo o de una medición (instrumento de medida: metro, balanza, cronómetro, termómetro, etc.); se divide a su vez en:

- <u>Variable cuantitativa Discreta</u>

Una variable se considera cuantitativa discreta cuando los valores que toma la variable provienen de un conteo. También

se dice que los valores que toma esta variable están asociados a los números enteros.

Ejemplos:

a. Y = Número de hijos por familia.
b. Z = Número de alumnos matriculados por cursos, en la Universidad Nacional Mayor de San Marcos en el II-Semestre 2003.
c. W =Número de accidentes de tránsitos registrados diariamente en el distrito de Miraflores, durante una semana cualquiera.
d. T = Número de artículos publicados en las revistas científicas de biología, durante los últimos 10 años.
e. P= Número de veces que frecuentan los usuarios durante la semana a la biblioteca Central de la UNMSM

- Variable cuantitativa continua

Una variable se considera cuantitativa continua, cuando los valores que toma la variable provienen de una medición; es decir, los valores que toma se encuentran en un intervalo. También se dice que los valores que toma esta variable están asociados a los números reales.

Ejemplos:

a. X = Montos (S/.) por ventas diarias por los grifos en el distrito de la Molina.
b. Y = Edad (años) de los alumnos matriculados en el Semestre Académico 2004-II en la Universidad Nacional Mayor de San Marcos.
c. Z = Tiempo de vida (horas) de baterías para automóviles.
d. Y = Estatura (mts) de los Cadetes de la EO-PNP.

Variables

| Tipo | VARIABLES CATEGÓRICAS | | | | VARIABLES NUMÉRICAS | | | |
|---|---|---|---|---|---|---|---|---|
| Naturaleza | CUALITATIVAS | | | | CUANTITATIVAS | | | |
| Escala | NOMINAL | | ORDINAL | | INTERVALO | | RAZÓN | |
| Ejemplos | Género | Estado Civil | Instrucción | Intensidad | Temperatura | Hora del día | Peso | Hijos |
| Valor Final | Masculino<br>Femenino | Soltero<br>Casado<br>Conviviente | Primaria<br>Secundaria<br>Superior | Leve<br>Moderado<br>Severo | -10°C<br>0°C<br>20°C | 00 Horas<br>10 Horas<br>20 Horas | 00.00Kg<br>10.24Kg<br>20.00Kg | Uno<br>Dos<br>Tres |
| Observaciones | **Dicotómicas:** Tienen solamente dos categorías<br>Ejemplos de Ordinal Dicotómica:<br>Nuevo - Continuador<br>Vivo – Fallecido<br>Sano – Enfermo<br>**Politómicas:** Tienen más de dos categorías. | | | | **Continuas:** Provienen de medir<br>Se pueden representar con números enteros o fraccionarios<br>Entre dos valores siempre existe un valor intermedio<br>**Discretas:** Provienen de contar<br>Solamente pueden ser representados con números enteros | | | |

Fuente: Apuntes de clase Supo (2020)

### 1.3.6. Muestra

Es una parte de la población en estudio o también se dice que es un subconjunto de unidades elementales de la población. Al tamaño de la muestra se le representa con la letra minúscula “n”. Se dice que una muestra es aleatoria o al azar, cuando todos los elementos de la población tienen la misma probabilidad de ser parte de la muestra.

Muestreo, es el procedimiento para obtener una muestra representativa de la población en estudio.

### 1.3.7. Parámetro

Es un valor fijo que caracteriza a la variable de interés en la población en estudio. Para calcular este valor se requiere conocer todos los datos de la variable en la población; por lo general este valor casi nunca se conoce el objetivo de investigador es estimarlo a través de una muestra aleatoria.

Los parámetros más utilizados son:

1. La media aritmética poblacional $(\mu)$
2. La mediana poblacional $(Me)$
3. La moda Poblacional $(Mo)$

4. La media Geométrica Poblacional $(\mu_G)$
5. La media Armónica Poblacional $(\mu_A)$
6. La variancia Poblacional $(\sigma^2)$
7. La desviación Estándar Poblacional $(\sigma)$
8. El coeficiente de variación Poblacional $(CV)$
9. La proporción Poblacional $(P)$
10. El rango poblacional $(R)$

1.3.8. Valor Estadístico o estimado.

Es un valor calculado con los datos de la muestra aleatoria y se supone estima al valor del parámetro de la variable en estudio en la población. Este valor cambia de una muestra a otra muestra. La fórmula que permite calcular el valor estadístico recibe el nombre de estimador o estadística.

Los valores estadísticos más utilizados son:

a. La media aritmética muestral $(\bar{x})$
b. La mediana muestral $(m_e)$
c. La moda muestral $(m_o)$
d. La media geométrica muestral $(\bar{x}_g)$
e. La media armónica muestral $(\bar{x}_a)$
f. La variancia muestral $(s^2)$
g. La desviación estándar muestral $(s)$
h. El coeficiente de variación muestral $(cv)$
i. La proporción muestral $(p)$
j. El rango muestral $(r)$

## 1.4. Ejercicios resueltos

Ejemplo-1

En Lima Metropolitana se toma una muestra aleatoria de 30 grifos que se dedican a la venta de combustible de un día determinado, se obtiene

que el monto por venta promedio es de S/. 12,600.50. Con esta información identificar: población, unidad elemental, variable en estudio, tipo de variable, un posible dato, tamaño de la muestra, parámetro y valor estadístico.

Solución:

| | | |
|---|---|---|
| Población | : | Todos los grifos de Lima Metropolitana, que se dedican a la venta de combustible. |
| Unidad elemental | : | Cada grifo de Lima Metropolitana que se dedica a la venta de combustible |
| Variable en estudio | : | Monto por venta(S/.) diaria de combustible |
| Tipo de variable | : | Variable cuantitativa continua |
| Dato u observación | : | S/12,400.30 |
| Tamaño de la muestra | : | 30 grifos |
| Parámetro | : | No se conoce el monto promedio por venta diaria en la población. |
| Valor estadístico | : | S/.12,600.50, venta promedio en la muestra. |

Ejemplo-2

De una población de 2,042 revistas científicas de biología, el 73.46% están escritas en idioma inglés. Una muestra aleatoria de 50 revistas arroja que el 72.0% están escritas en inglés; el promedio de artículos por revistas es de 12.4. Con esta información identificar: población, unidad elemental, variables en estudios, tipos de variables, posibles datos, tamaño de la muestra, parámetro y estimado o valor estadístico.

Solución:

| | | |
|---|---|---|
| Población | : | Todas las revistas científicas de biología N = 2,042. |
| Unidad elemental | : | Cada revista científica de biología. |
| Tamaño de la muestra | : | 50 revistas científicas de biología |

| Variables en estudio | Tipos de variables | Datos u observaciones | Parámetro | Valor estadístico o estimado |
|---|---|---|---|---|
| Idioma de las revistas científicas | Cualitativa nominal | Español, ingles, portugués, etc. | 73.46%, están en idioma ingles | 72.0%, están en idioma ingles |
| Números de artículos por revistas | Cuantitativa discreta | 12, 16, 11, 9, etc. | Se desconoce | 12.4 promedio del número de artículos por revistas |

1.5. Ejercicios propuestos

En cada uno de los siguientes enunciados identificar de ser posible: la población, unidad elemental, variables en estudio, tipos de variables, tamaño de la muestra, valores estadísticos, parámetros y un ejemplo de posibles datos en cada caso:

1. Una muestra al azar de 50 recién nacidos en la maternidad de Lima arroja un peso promedio de 3.50 Kg; el 46% son varones y 70% tuvieron parto normal.
2. Una muestra aleatoria de 550 peruanos mayores de 18 años arroja la siguiente información: la edad promedio es 32 años;20% tienen grado de instrucción superior y 30% está de acuerdo con las medidas económicas anunciadas por el gobierno hace unos días.
3. Para analizar la posibilidad de lanzar al mercado una nueva marca de cigarrillos, se efectuó una encuesta entre fumadores que transitaron por la quinta cuadra de la avenida Canadá entre las 5 p.m. y 8 p.m. del día 14 de febrero; para ello se seleccionó en forma aleatoria a 80 fumadores, obteniéndose los siguientes resultados: el 25 % prefiere los cigarrillos importados que los nacionales y el precio promedio que pagan al comprar una cajetilla es de S/. 3.40.
4. Una muestra aleatoria de 100 microbuses de transporte urbano en Lima metropolitana a las 12.00 horas de un día cualquiera, arroja que el promedio de pasajeros dentro del microbús es de 18.4 personas; el 75% de los microbuses tienen más de 10 años de antigüedad.
5. Una muestra aleatoria de 50 frascos conteniendo mermelada arroja un peso promedio de 45.3 grs. Y una desviación estándar de 5.436 grs. Si la producción diaria de la planta es de 5,000

frascos con un peso promedio de 45.5 grs y una desviación estándar de 5.5 grs.

6. El porcentaje de varones de un centro educativo es de 62.5%; una muestra aleatoria de 50 alumnos arroja que 58% son varones; la edad promedio es de 15.4 años; el promedio de hermanos es de 2.3 y el 60% tienen casa propia.
7. Una muestra aleatoria de 70 usuarios que concurren a un centro de información arroja que el 34.3% son del sexo masculino. De experiencias pasadas se sabe que el porcentaje de varones que concurren fue de 34.2%.
8. Una muestra aleatoria de 20 revistas científicas sobre Biología arroja que el promedio de artículos publicados es de 13.3; el 70% están escritos en inglés; el 15% son de nacionalidad Argentina.
9. De una población de 1,200 actas de sustentación en Matemáticas Pura, el 72% obtuvo el calificativo de bueno al sustentarlo. Una muestra aleatoria de 25 actas arroja que el 76% de actas obtuvieron el calificativo de bueno.
10. Una muestra al azar de 65 Cadetes de la Escuela de Oficiales de la Policía Nacional del Perú arroja los siguientes resultados: El 78.4% son de sexo masculino; el 32% cursa el 2do año académico; el 80% provienen de colegios particulares; el gasto promedio semanal en los cafetines es de 30.4 nuevos soles; el promedio de hermanos es de 2.4; la edad promedio es de 23.4 años y su estatura promedio es de 1.69 mts.
11. Una muestra al azar de 0 estudiante de la Facultad de letras y ciencias Humanas de la UNMSM arroja la siguiente información: el 30% tienen edades superiores a los 30 años; el 52% son de sexo masculino; el promedio de créditos matriculados es de 21.4; la edad promedio es de 23.3 años; el promedio de cursos llevados es de 5.2; el 32% de sus profesores tienen grado de Magister.

## 1.6. ETAPAS DEL MÉTODO ESTADÍSTICO

La estadística, proporciona métodos y procedimientos que permiten analizar datos de variables de interés, utilizando para tal fin una muestra elegida al azar de la población en estudio; con el objetivo de tomar decisiones en dicha población. Para su aplicación, se debe tener en cuenta las siguientes etapas:

### 1.6.1. Planteamiento del problema.

Debe plantearse con precisión y claridad la necesidad que se desea conocer, las causas de la investigación.

1.6.2. Objetivo

Conjunto de acciones o tareas de las necesidades alcanzadas, debe plantearse un objetivo general y como mínimo tres objetivos específicos.

1.6.3. Determinación de la unidad de investigación

Corresponde al elemento de la población, que puede ser persona, animal u objeto, de la cual se definen variable de interés para el investigador.

1.6.4. Formulación de hipótesis

Son supuestos que se plantea en la población con respecto a la distribución o parámetro de la variable de interés, estos pueden ser rechazado o no, utilizando para tal propósito una muestra aleatoria.

1.6.5. Selección del método de investigación

La investigación puede ser exploratorio, descriptivo, correlacional, etc.

1.6.6. Recopilación de datos.

Corresponde a la recolección de datos. Esta puede ser por observación directa, por encuesta, por publicaciones y/o fuentes externas confiables. El éxito de toda investigación depende en gran parte de la consistencia de los datos obtenidos.

1.6.7. Procesamiento de los datos

Corresponde a obtener tablas, gráficos o valores estadísticos, en la actualidad se usan software tales: Excel, Minitab, SPSS, SAS, etc.

1.6.8. Publicación

Se consigna los resultados de la investigación, previa revisión de los datos procesados.

1.6.9. Generalización e inferencia estadística.

Las conclusiones de la investigación son generalizadas a la población de donde se obtuvo la muestra aleatoria.

## 1.7. SUMATORIAS SIMPLES, DOBLES Y NOTACIÓN PUNTUAL

### 1.7.1. SUMATORIAS SIMPLES

Sea X, una variable cuantitativa en estudio que toma valores $x_1, x_2, ..., x_n$, la suma de los valores $x_1 + x_2 +,...,+x_n$, se representa mediante el operador matemático $\sum_{i=1}^{n} x_i$; que significa sumar todos los valores $x_i$, para "$i$" que toma valores enteros positivos consecutivos desde 1 hasta n; es decir:

$$x_1 + x_2 +,...,+x_n = \sum_{i=1}^{n} x_i$$

Donde:

$x_i =$ Valor que toma la variable $X$, en la i-ésima observación

Ejemplos:

1. Desarrollar de las siguientes sumatorias

$$\sum_{i=1}^{5} x_{2i+1} = x_3 + x_5 + ... + x_{11}$$

$$\sum_{i=1}^{6} x_{2i}^2 = x_2^2 + x_4^2 + ... + x_{12}^2$$

$$\sum_{i=1}^{4} \frac{2x_{i+1}}{y_{2i+1}} = \frac{2x_2}{y_3} + \frac{2x_3}{y_5} + ... + \frac{2x_5}{y_9}$$

$$\sum_{i=1}^{5} (x_i - 7)^2 = (x_1 - 7)^2 + (x_2 - 7)^2 + ... + (x_5 - 7)^2$$

2. Supongamos que una variable $X$, toma los siguientes valores:

$x_1 = 7 \quad x_2 = 9 \quad x_3 = 5 \quad x_4 = 6$

Calcular: $\sum_{i=1}^{4} \frac{2x_i^2+3}{x_i-2}$

$$\sum_{i=1}^{4} \frac{2x_i^2+3}{x_i-2} = \frac{2x_1^2+3}{x_1-2} + \frac{2x_2^2+3}{x_2-2} + \frac{2x_3^2+3}{x_3-2} + \frac{2x_4^2+3}{x_4-2}$$

$$\sum_{i=1}^{4} \frac{2x_i^2+3}{x_i-2} = \frac{2(7)^2+3}{7-2} + \frac{2(9)^2+3}{9-2} + \frac{2(5)^2+3}{5-2} + \frac{2(6)^2+3}{6-2}$$

$$\sum_{i=1}^{4} \frac{2x_i^2+3}{x_i-2} = 20.2 + 23.57 + 17.67 + 18.75 = 80.19$$

3. Supongamos que una variable $X$ toma los siguientes valores:

$$x_1 = -2 \quad x_2 = 1 \quad x_3 = -4 \quad x_4 = 3$$

Calcular: $\sum_{i=1}^{4} (x_i + 2)^2$

$$\sum_{i=1}^{4} (x_i + 2)^2 = (x_1 + 2)^2 + (x_2 + 2)^2 + (x_3 + 2)^2 + (x_3 + 2)^2$$

$$\sum_{i=1}^{4} (x_i + 2)^2 = (-2 + 2)^2 + (1 + 2)^2 + (-4 + 2)^2 + (3 + 2)^2 = 38$$

Calcular: $\sum_{i=1}^{4} (2x_i + 3)(x_i - 7)$

$$\sum_{i=1}^{4} (2x_i + 3)(x_i - 7) = (2x_1 + 3)(x_1 - 7) + (2x_2 + 3)(x_2 - 7) ... + (2x_4 + 3)(x_4 - 7)$$

$$\sum_{i=1}^{4}(2x_i+3)(x_i-7)=(2(-2)+3)((-2)-7)+(2(1)+3)((1)-7)...+(2(3)+3)((3)-7)$$

$$\sum_{i=1}^{4}(2x_i+3)(x_i-7)=-2$$

De otra forma

En una tabla establecer varias columnas, la primera para el subíndice que toma la variable, la segunda para los valores que toma la variable $X$ , la tercera para el primer factor, la cuarta para el segundo factor y la quinta para el producto de los dos factores, luego para la tercera y cuarta columna se calculan los valores numéricos para cada valor de la variable y para la quinta el producto de los valores obtenidos en los pasos anteriores, y la suma de estos valores es el resultado pedido.

| $i$ | $x_i$ | $2x_i+3$ | $x_i-7$ | $(2x_i+3)(x_i-7)$ |
|---|---|---|---|---|
| 1 | -2 | -1 | -9 | 9 |
| 2 | 1 | 5 | -6 | -30 |
| 3 | -4 | -5 | -11 | 55 |
| 4 | 3 | 9 | -4 | -36 |
| | | | | -2 |

Es decir: $\sum_{i=1}^{4}(2x_i+3)(x_i-7)=9-30+55-36=-2$

**PROPIEDADES DE LAS SUMATORIAS SIMPLES**

Sea la sumatoria:

$$\sum_{i=a}^{b} x_i \,,$$

Siendo a y b números enteros $(a \le b)$ , y el subíndice "$i$" que toma valores enteros consecutivos desde “a” hasta “b”, donde: el

valor de “a” es el límite inferior y “b” es el límite superior de la sumatoria. Teniendo en cuenta la definición de sumatoria, se tienen las siguientes propiedades matemáticas:

i. El número de términos o sumandos de una sumatoria es igual al limite superior menos el limite inferior más la unidad; es decir,

$$N^{o}Términos = b - a + 1$$

ii. La sumatoria de un valor constante, es igual a la constante multiplicada por el número de términos de la sumatoria; es decir, sea $x_i = m \quad \forall i = a,\ a+1,\ a+2, ..., b$, entonces:

$$\sum_{i=a}^{b} m = m(b - a + 1)$$

Siendo: $m$ la constante

iii. La sumatoria de una constante multiplicada por una variable, es igual a la constante por la sumatoria de la variable; es decir:

$$\sum_{i=a}^{b} mx_i = m\sum_{i=a}^{b} x_i$$

Siendo: $m$ la constante

iv. Sea la combinación lineal $x_i = ny_i \pm pz_i \pm q$, la sumatoria de los $x_i$ para $i = a,\ a+1,\ a+2, ..., b.$, es igual a:

$$\sum_{i=a}^{b} x_i = \sum_{i=a}^{b} (ny_i \pm pz_i \pm q) = n\sum_{i=a}^{b} y_i \pm p\sum_{i=a}^{b} z_i \pm q(b - a + 1)$$

Siendo: $n, p, q$ constantes

### 1.7.2. SUMATORIAS DOBLES

Sea $X$ , una variable en estudio, cuyo valor es el resultado de observar simultáneamente dos características a una unidad elemental, valor que toma doble subíndice $x_{ij}$, representa el valor que toma la variable $X$ en la $i = ésima\ fila$ (primera variable) y la $j = ésima\ columna$ (segunda variable), por lo tanto si se tiene "a" filas y "b" columnas, la suma total de los valores de la variable $X$ se puede representar por la doble sumatoria; es decir,

$$\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}$$

Ejemplo: desarrollar

$$\sum_{i=1}^{2}\left(\sum_{j=1}^{3} x_{ij}\right) = \underbrace{\sum_{j=1}^{3} x_{1j} + \sum_{j=1}^{3} x_{2j}}_{desarrollando\ el\ subindice\ "i"} = \underbrace{x_{11}+x_{12}+x_{13} + x_{21}+x_{22}+x_{23}}_{desarrollando\ el\ subindice\ "j"}$$

**PROPIEDADES DE LAS SUMATORIAS DOBLES**

Sea la sumatoria doble:

$$\sum_{i=a}^{b}\sum_{j=c}^{d} x_{ij}$$

Siendo $a, b, c, d$ números enteros positivos, y los subíndices "i", "j" toma valores consecutivos desde "a" hasta "b" y desde "c" hasta "d" respectivamente; los valores "a" y "c" son los límites inferiores, y "b" y "d" son los límites superiores de la doble sumatoria.

Teniendo en cuenta la definición de sumatoria doble, se tienen las siguientes propiedades matemáticas:

i. El número de términos de una doble sumatoria es igual al producto del número de términos de la primera sumatoria

con el número de términos de la segunda sumatoria; es decir:

$$N^{o}\ Términos = (b - a + 1)(d - c + 1)$$

ii. La doble sumatoria de una constante, es igual a la constante multiplicada por el número de términos de la sumatoria; es decir, sea $x_{ij} = m, \quad \forall ij$, entonces:

$$\sum_{i=a}^{b}\sum_{j=c}^{d} m = m(b - a + 1)(d - c + 1)$$

Siendo: $m$ la constante

iii. La doble sumatoria de una constante multiplicada por una variable, es igual a la constante por la doble sumatoria de la variable; es decir:

$$\sum_{i=a}^{b}\sum_{j=c}^{d} m x_{ij} = m \sum_{i=a}^{b}\sum_{j=c}^{d} x_{ij}$$

Siendo: $m$ la constante

iv. Sea la combinación lineal $x_{ij} = n y_{ij} \pm p z_{ij} \pm q$, la doble sumatoria de los $x_{ij}$, es igual a:

$$\sum_{i=a}^{b}\sum_{j=c}^{d} x_{ij} = \sum_{i=a}^{b}\sum_{j=c}^{d} (n y_{ij} \pm p z_{ij} \pm q) = n\sum_{i=a}^{b}\sum_{j=c}^{d} y_{ij} \pm p\sum_{i=a}^{b}\sum_{j=c}^{d} z_{ij} \pm q(b - a + 1)(d - c + 1)$$

Siendo: $n, p, q$ constantes

### 1.7.3. NOTACIÓN PUNTUAL PARA UNA SUMATORIA DOBLE

Sea la tabla de doble entrada (conocida como tabla de contingencia)

| "$J$" / "$i$" | 1 | 2 | 3 | … | $k$ | Total |
|---|---|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{12}$ | $x_{13}$ | … | $x_{1k}$ | $x_{1.}$ |
| 2 | $x_{21}$ | $x_{22}$ | $x_{23}$ | … | $x_{2k}$ | $x_{2.}$ |
| 3 | $x_{31}$ | $x_{32}$ | $x_{33}$ | … | $x_{3k}$ | $x_{3.}$ |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . |
| $r$ | $x_{r1}$ | $x_{r2}$ | $x_{r3}$ | … | $x_{rk}$ | $x_{r.}$ |
| Total | $x_{.1}$ | $x_{.2}$ | $x_{.3}$ | … | $x_{.k}$ | $x_{..}$ |

De la tabla la suma de los elementos de la primera columna

$x_{11} + x_{21} + x_{31} + ... + x_{r1} = \sum_{i=1}^{r} x_{i1}$ ; al valor de la sumatoria $\sum_{i=1}^{r} x_{i1}$ se le representa mediante la siguiente notación puntual $x_{.1}$ ; es decir,

$$\sum_{i=1}^{r} x_{i1} = x_{.1}$$

$x_{.1}$ = corresponde a la sumatoria simple de todos los elementos $x_{i1}$, para "$i$" que toma valores desde 1 hasta r, de la primera columna.

De la tabla la suma de todos los elementos de la 1ra fila

$x_{11} + x_{12} + x_{13} + ... + x_{1k} = \sum_{j=1}^{k} x_{1j} = x_{1.}$ (Corresponde a la sumatoria simple de todos los elementos $x_{1j}$, para "$j$" que toma valores desde 1 hasta k, de la primera fila)

También de la tabla

$$\sum_{i=1}^{r}\sum_{j=1}^{k} x_{ij} = \sum_{i=1}^{r} x_{i.} = x_{..}$$

## 1.8. Ejercicios resueltos

*1.* Sean $\sum_{i=1}^{50} x_i = 20$ *y* $\sum_{i=1}^{50} x_i^2 = 100$ *hallar:* $\sum_{i=1}^{50}(3x_i - 4)^2$

Solución

$$\sum_{i=1}^{50}(3x_i - 4)^2 = \sum_{i=1}^{50}(9x_i^2 - 24x_i + 16) = \sum_{i=1}^{50} 9x_i^2 - \sum_{i=1}^{50} 24x_i + \sum_{i=1}^{50} 16 =$$

$$= 9\sum_{i=1}^{50} x_i^2 - 24\sum_{i=1}^{50} x_i + 16(50 - 1 + 1) = 9(100) - 24(20) + 16(50) = 1220$$

2. Simplificar

$$\sum_{i=1}^{n}(x_i^2 + 5x_i + 5)^2 - \sum_{i=1}^{n}(x_i + 1)(x_i + 2)(x_i + 3)(x_i + 4)$$

Solución

$$\sum_{i=1}^{n}(x_i^2 + 5x_i + 5)^2 - \sum_{i=1}^{n}(x_i + 1)(x_i + 2)(x_i + 3)(x_i + 4)$$

$$\sum_{i=1}^{n}(x_i^2 + 5x_i + 5)^2 - \sum_{i=1}^{n}(x_i + 1)(x_i + 4)(x_i + 2)(x_i + 3)$$

$$\sum_{i=1}^{n}(x_i^2 + 5x_i + 5)^2 - \sum_{i=1}^{n}(x_i^2 + 5x_i + 4)(x_i^2 + 5x_i + 6) \quad (1)$$

Supongamos que $y_i = x_i^2 + 5x_i$, luego reemplazando en (1) se tiene

$$\sum_{i=1}^{n}(y_i+5)^2 - \sum_{i=1}^{n}(y_i+4)(y_i+6) = \sum_{i=1}^{n}\left\{(y_i+5)^2-(y_i+4)(y_i+6)\right\}$$

$$\sum_{i=1}^{n}\left\{(y_i^2+10y_i+25)-(y_i^2+10y_i+24)\right\} = \sum_{i=1}^{n}\{1\} = 1(n-1+1) = n$$

3. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-x_{i.}-x_{.j}+x_{..})$$

Solución

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-x_{i.}-x_{.j}+x_{..}) =$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}-\sum_{i=1}^{a}\sum_{j=1}^{b}x_{i.}-\sum_{i=1}^{a}\sum_{j=1}^{b}x_{.j}+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{..}$$

$$= x_{..}-b\sum x_{i.}-a\sum x_{.j}+abx_{..}$$

$$= x_{..}-bx_{..}-ax_{..}+abx_{..} = (1-a)(1-b)x_{..}$$

4. Sean: $\sum_{i=1}^{n}x_i^2 = 100$ $y$ $\sum_{i=1}^{n}x_i = 5$, Calcular el valor de

$$\frac{\sum_{i=1}^{n}(x_i-10)^2}{n}$$

Desarrollando el numerador de la sumatoria y aplicando las propiedades:

$$\sum_{i=1}^{n}(x_i-10)^2=\sum_{i=1}^{n}(x_i^2-20x_i+100)=\sum_{i=1}^{n}x_i^2-20\sum_{i=1}^{n}x_i+100n=100-20(5)+$$

Luego se tiene que: $\frac{100n}{n}=100$

5. De la tabla

| $i$ | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 4 | 5 |
| 2 | 5 | 6 |
| 3 | 2 | -4 |
| 4 | 3 | 5 |
| 5 | 2 | 7 |

Calcular el valor de : $\sum_{i=1}^{5} 2(7x_i^2-5y_i^3)^2$

$$\sum_{i=1}^{5} 2(7x_i^2-5y_i^3)^2=2\sum_{i=1}^{5}(7x_i^2-5y_i^3)^2$$

De la tabla

| $i$ | $x_i$ | $y_i$ | $7x_i^2$ | $5y_i^3$ | $7x_i^2-5y_i^3$ | $(7x_i^2-5y_i^3)^2$ |
|---|---|---|---|---|---|---|
| 1 | 4 | 5 | 112 | 625 | -513 | 263169 |
| 2 | 5 | 6 | 175 | 1080 | -905 | 819025 |
| 3 | 2 | -4 | 28 | -320 | 348 | 121104 |
| 4 | 3 | 5 | 63 | 625 | -562 | 315844 |
| 5 | 2 | 7 | 28 | 1715 | -1687 | 2845969 |
| Suma | | | | | | 4365111 |

$$\sum_{i=1}^{5}2(7x_i^2-5y_i^3)^2=2(4'365{,}111)=8'730{,}222$$

## 1.9. Ejercicios Propuestos:

De la tabla que se presenta a continuación:

| $k$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x_k$ | 2 | 1 | 0 | 3 | 1 |

Calcular:

1. $$\sum_{k=1}^{5}(5+3x_k)^2$$

2. $$\sum_{k=1}^{5}(x_k^2 - x_k + 1)^2$$

3 $$\sum_{k=1}^{5}\left\{(x_k-\bar{x})^2-(x_k+\bar{x})^2\right\}$$ donde $$\bar{x}=\frac{\sum_{k=1}^{5} x_k}{5}$$

4. $$\sum_{k=1}^{5}\left\{(5x_k-4)(5x_k+4)\right\}$$

5. $$\left\{\sum_{k=1}^{5}(x_k-1)^3\right\}^2$$

6. $$\sum_{k=1}^{5}(x_k-5)\sum_{k=1}^{5}(x_k+5)$$

7. De la información dada en la tabla, si a cada valor se le incrementa en un 40% más 3 unidades. Hallar la suma de los cuadrados de los nuevos valores.

Dada las siguientes expresiones simplificar:

8. $\sum_{i=1}^{n}(x_i - \bar{x})^2$ , siendo $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

9 $\sum_{i=1}^{n}(x_i - \bar{x})x_i - \sum_{i=1}^{n}(x_i - \bar{x})^2$ , siendo $\bar{x} = \frac{\sum_{i=1} x_i}{n}$

*10.* $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + n\bar{x}\bar{y}$

Donde: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

11. Sea $y_i = ax_i \pm c$ para $i = 1,2,..., n$,

calcular el valor de $\sum_{i=1}^{n}(y_i - \bar{y})^2$

*12.* Sea $\sum_{i=1}^{5}(x_i - 3)^2 = 18$ donde $\sum_{i=1}^{5} x_i = 7$

Calcular el valor de $\sum_{i=1}^{5}(x_i + 3)^2$

13. Sea $\sum_{i=1}^{5}(x_i + 5)^2 = 200$ , calcular el valor de

$\sum_{i=1}^{5}(x_i - \bar{x})^2$ , donde $\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = 1$

Sean X e Y dos variables que toman valores tal como se presenta en la tabla.

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x_i$ | 1.65 | 1.70 | 1.75 | 1.80 | 1.85 |
| $y_i$ | 60 | 65 | 70 | 75 | 80 |

Calcular:

14. $\sum_{i=1}^{5}(2x_i + y_i)^2$

15. $\sum_{i=1}^{5}\left\{2(x_i+\bar{x})(y_i-\bar{y})-11\right\}$, siendo $\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}$ $\bar{y}=\frac{\sum_{i=1}^{5}y_i}{5}$

16. $\frac{\sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{5}(x_i-\bar{x})x_i}$, siendo $\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}$ $\bar{y}=\frac{\sum_{i=1}^{5}y_i}{5}$

17. $\sum_{i=1}^{5}\frac{\sqrt{x_i+y_i}}{x_i y_i}$

18. $\sum_{i=1}^{5}(y_i-\bar{y})^2-\sum_{i=1}^{5}(x_i-\bar{x})^2$, siendo $\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}$ $\bar{y}=\frac{\sum_{i=1}^{5}y_i}{5}$

19. Calcular el valor de "M" en: $\sum_{i=1}^{5}(x_i-M)^2=10$, Siendo:

$\sum_{i=1}^{5}x_i=15$ y $\sum_{i=1}^{5}x_i^2=55$

Dada la siguiente tabla de doble entrada con valores $x_{ij}$

| i \ j | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 2 | 1 | 0 | 0 |
| 2 | 4 | 1 | 2 | 3 |
| 3 | 0 | 2 | 0 | 1 |

Calcular las siguientes sumatorias:

20. $\sum_{i=1}^{3}\sum_{j=1}^{4}(x_{ij} - \bar{x}_{..})^2$

21. $\sum_{i=1}^{3}(x_{i.}^2 + \sum_{j=1}^{4} x_{.j}^2 + x_{..}^2)$

22. $\sum_{j=1}^{4}(x_{.i} + 2)^2$

23. $\sum_{i=1}^{3}(\sum_{j=1}^{4} x_{ij})^2$

24. $\sum_{i=1}^{3}\sum_{j=1}^{4}(x_{.j}^2 + x_{i}^2 - x_{..}^2 + 24x_{.j})$

25. Sea la siguiente ecuación $\sum_{i=1}^{n}(x_i + 2) = 5n$, Hallar en valor de:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

26. Hallar la suma de los coeficientes al desarrollar:

$$\sum_{i=1}^{5}(3x_i + 5y_i)^{12}$$

27. Simplificar

$$\sum_{i=1}^{n+1} 2^{i-1} + \sum_{i=1}^{n} 2^{i} - 2\sum_{i=0}^{n} 2^{n-i}$$

En un asiento minero de Cerro de Pasco, para un mes determinado, la administración hace la distribución del personal de ingenieros por turno y modalidad de trabajo, tal como se presenta en la tabla:

| Modalidades / Turnos | Tajo Abierto =1 | Socavón = 2 | Seguridad = 3 |
|---|---|---|---|
| Turno =1 | 16 | 20 | 2 |
| Turno =2 | 30 | 40 | 2 |
| Turno =3 | 0 | 70 | 4 |

Donde X ij = corresponde al número de ingenieros para el i-ésima turno en la j-ésima modalidad de trabajo. Para cada caso expresar por sumatoria y calcular su valor:

28. Total de ingenieros que laboran en el 1er. Y 3er turno, para todas las modalidades.
29. Total de ingenieros que laboran en el 1er. Y 3er. Turno, para las modalidades de trabajo: tajo abierto y seguridad.
30. Si los ingenieros del 1er. Turno ganan “b-a” nuevos soles, para los del 2do. Turno “b” nuevos soles y para los del tercero “b+a” nuevos soles. Cual es el total a pagar por la administración en ese mes.

31. Simplificar

$$\sum_{i=1}^{n}(x_i^2 + 13x_i + 41)^2 - \sum_{i=1}^{n}(x_i + 5)(x_i + 6)(x_i + 7)(x_i + 8)$$

32. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_i + y_j - \bar{x} - \bar{y})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}(x_i^2 + y_j^2 - \bar{x}x_i - \bar{y}y_j)$$

Donde : $\bar{x} = \dfrac{\sum_{i=1}^{a} x_i}{a}$ y $\bar{y} = \dfrac{\sum_{j=1}^{b} y_j}{b}$

33. Simplificar

$$\sum_{i=1}^{n}(x_i^4 + 5x_i^2 + 5)^2 - \sum_{i=1}^{n}(x_i^2 + 1)(x_i^2 + 2)(x_i^2 + 3)(x_i^2 + 4)$$

34. Simplificar

$$\sum_{i=1}^{n+1}(x_i - \bar{x})x_i - \sum_{i=1}^{n+1}(x_i - \bar{x})^2$$

35. Simplificar

$$\sum_{i=1}^{n}(x_i + a)^2 - \sum_{i=1}^{n}(x_i - a)^2$$

36. Simplificar

$$\sum_{i=1}^{2}\sum_{j=1}^{3}(5x_{i.} + 5x_{.j} - 2x_{1.} - 2x_{2.})$$

37. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{..})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{.j})^2 + a\sum_{j=1}^{b}(\bar{x}_{.j} - \bar{x}_{..})^2$$

38. Simplificar

$$\sum_{i=1}^{a+1}\sum_{j=1}^{b+1}x_{ij} - \sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}$$

39. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - x_{i.})x_{.j} - \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - x_{.j})x_{i.}$$

40. Sea $x_1, x_2, .... x_{100}$, una muestra de 100 datos, cuya suma de sus valores es igual a 400. Si a cada dato se le incrementa en 5% de su valor más la unidad, calcular la nueva suma.

41. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - x_{..})x_{ij} - \sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}^2 - \frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\right)^2$$

42. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i.} + \bar{x}_{.j} + x_{ij} + x_{.j} + x_{i.})$$

43. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i.})^2 + \frac{1}{b}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}x_{ij}\right)^2$$

44. Simplificar

$$\sum_{i=1}^{11}(x_i + 1)x_i - \sum_{i=1}^{12}(x_i + 4)(x_i - 3) + \sum_{12}^{12}(x_i + 12)(x_i - 11)$$

45. Simplificar

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..})$$

46. Sea $w_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ donde $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Calcular el valor de : $\sum_{i=1}^{n} w_i x_i$

La Empresa de transporte San Francisco S.A, cuenta con 4 vehículos cuya ruta comprende los distritos del Callao con la Molina. Se define a la variable en estudio Xij= Numero de boletos entregados a los usuarios en el i-ésimo vehículo para el j-ésimo turno, tal como se presenta en el siguiente cuadro:

| Vehículos / Turnos | Vehículo=1 | Vehículo=2 | Vehículo=3 | Vehículo=4 |
|---|---|---|---|---|
| Turno=1 | 400 | 500 | 600 | 600 |
| Turno=2 | 300 | 400 | 400 | 300 |
| Turno=3 | 400 | 400 | 400 | 400 |

Expresar mediante sumatorias y calcular su valor

46. El Total de boletos expedidos a los usuarios del 1er y 3er turno para todos los vehículos de la empresa.
47. Si el precio para el tercer turno es el doble del precio del 2do y el precio del primer turno es la mitad del 2do turno, siendo el precio del 2do turno igual a “q” nuevos soles.
48. El total de boletos expedidos por los tres turnos para el 1ro y 4to vehículo.

49. Simplificar

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})z_i - n\bar{z}\sum_{j=1}^{n}(x_j - \bar{x})y_j ,$$ siendo:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \bar{z} = \frac{\sum_{i=1}^{n} z_i}{n}$$

50. Dada la siguiente tabla con valores $x_{ij}$ = Valor que toma la variable X en la i-ésima fila y j-ésima columna.

| $i$ \ $j$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 2 | 1 | 2 |
| 2 | 3 | 2 | 0 |
| 3 | 2 | 2 | 3 |
| 4 | 4 | 1 | 2 |

Calcular:

a. $\sum_{i=1}^{4}(x_{i.} - \bar{x}_{..})^2$

b. $\sum_{i=1}^{4} \frac{\sum_{j=1}^{3} x_{ij}}{\sum_{j=1}^{3} x_{j,4-j}}$

c. $\sum_{i=1}^{4}\sum_{j=1}^{4}(x_{i.} - \bar{x}_{.j})^2$

51. Expresar mediante sumatorias las siguientes sucesiones de términos:

a. $x_{1j}^{1} + x_{2j}^{2} + x_{3j}^{3} + \ldots + x_{kj}^{k}$

b. $x_{1,2}^{1} + x_{2,3}^{2} + x_{3,4}^{3} + \ldots + x_{k,k+1}^{k}$

c. $\frac{x_1}{x_2} + \frac{x_2}{x_3} + \frac{x_3}{x_4} + \ldots + \frac{x_k}{x_{k+1}}$

d. $(x_1 + y_2)^1 + (x_2 + y_3)^4 + (x_3 + y_4)^9 + \ldots (x_k + y_{k+1})^{k^2}$

52. Sea la tabla de datos:

| $i$ | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 5 | 1 |
| 2 | 4 | 2 |
| 3 | 3 | 3 |
| 4 | 2 | 4 |
| 5 | 1 | 5 |

**Calcular:** $\sum (x_i^2 + 2x_i y_i + y_i^2)^2$

53. Sea la tabla de datos:

| $i$ | $x_i$ | $y_i$ | $z_i$ |
|---|---|---|---|
| 1 | 2 | 3 | 3 |
| 2 | 1 | 2 | 2 |
| 3 | 1 | 3 | 4 |
| 4 | 2 | 2 | 1 |
| 5 | 1 | 2 | 2 |

**Calcular:**

**a)** $\sum_{i=1}^{5} (x_i + y_i)(x_i + z_i)(y_i + z_i)$

**a)** $\sum_{i=1}^{5} (x_i + 2)(y_i + 3)(z_i - 5)$

**b)** $\sum_{i=1}^{5} (4x_i + 3y_i - 5z_i)(x_i + 2y_i - 3z_i)$

**c)** $\sum_{i=1}^{5} \left[x_i(x_i + 2y_i) + y_i(y_i + 2z_i) + z_i(z_i + 2x_i)\right]$

**d)**

54. Sean las sumatorias:

$\sum_{i=1}^{20} x_i = 274$ $\sum_{i=1}^{20} x_i^2 = 3834$

**Hallar:**

$$a).\quad \sum_{i=1}^{20}\left[5(x_i - 2)^2 - 2(3x_i - 1)^2\right]$$

$$b).\quad \sum_{i=1}^{20}\left[x_i(7x_i + 2) - 3x_i(4x_i + 1)\right]$$

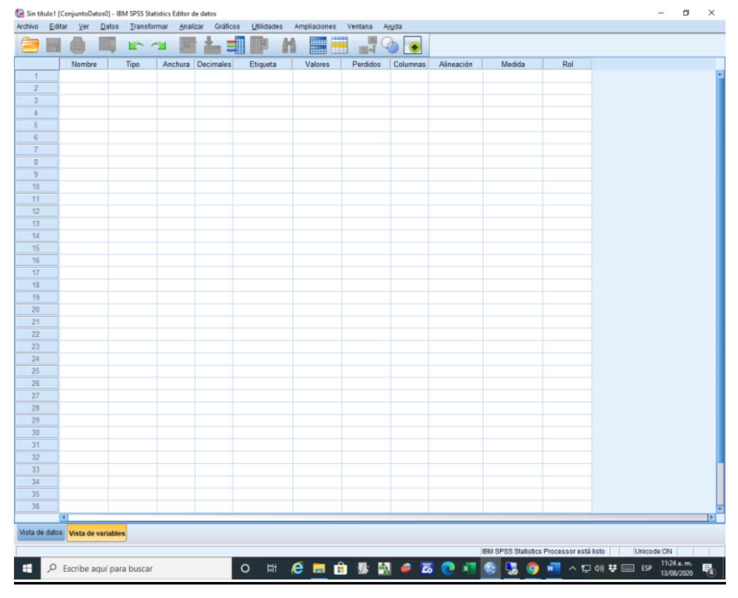
$$c).\quad \sum_{i=1}^{20}12\left[(2x_i + 4)(3x_i - 5)\right]$$

$$d).\quad \sum_{i=1}^{20}10(3x_i + 5)^2$$

$$e).\quad \sum_{i=1}^{20}(2x_i + 5)(2x_i - 5)$$

<u>Explorando el SPSS</u>

Al activar el SPSS mostrara una ventana con las opciones de vista de variables y vista de datos, que se puede seleccionar en la parte inferior izquierda:




Antes de empezar a ingresar los datos, lo que se tiene que hacer es definir las variables; la vista de variables presenta las siguientes opciones:

**Nombre**: Nombre abreviado de la variable, no acepta caracteres $, #; no permite dejar espacios en blanco.

**Tipo**: Para indicar el formato de la variable: numérico, cadena.

**Anchura**: Para determinar la cantidad de caracteres que tendrá un dato.

**Decimales**: Para datos de tipo numérico o moneda.

**Etiqueta**: Nombre completo de la variable.

**Valores**: Para el caso de variables cualitativas se puede definir a que categoría corresponde cada valor.

**Perdidos**: Indica que valores son codificados como datos perdidos o excluidos del análisis.

**Columna**: Define como será la alineación de la columna y se visualizará en la ventana de ingreso de datos.

**Alineación**: Define como será la alineación de los datos en la ventana de ingreso de datos.

**Medida**: El SPSS distingue tres tipos: escala, nominal y ordinal.

vimed16.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

| | Nombre | Tipo | Anchura | Decimales | Etiqueta | Valores | Perdidos | Columnas | Alineación | Medida | Rol |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Profesor | Numérico | 8 | 0 | Profesor | {1, Alberto}... | Ninguna | 8 | Derecha | Nominal | Entrada |
| 2 | Alumnos | Numérico | 8 | 0 | Alumnos | Ninguna | Ninguna | 8 | Derecha | Nominal | Entrada |
| 3 | Exámenes | Numérico | 8 | 0 | Exámenes | Ninguna | Ninguna | 8 | Derecha | Nominal | Entrada |
| 4 | Resultados | Numérico | 8 | 2 | Resultados | Ninguna | Ninguna | 8 | Derecha | Escala | Entrada |
| 5 | | | | | | | | | | | |
| 6 | | | | | | | | | | | |
| 7 | | | | | | | | | | | |
| 8 | | | | | | | | | | | |
| 9 | | | | | | | | | | | |
| 10 | | | | | | | | | | | |
| 11 | | | | | | | | | | | |
| 12 | | | | | | | | | | | |
| 13 | | | | | | | | | | | |
| 14 | | | | | | | | | | | |
| 15 | | | | | | | | | | | |
| 16 | | | | | | | | | | | |
| 17 | | | | | | | | | | | |
| 18 | | | | | | | | | | | |
| 19 | | | | | | | | | | | |
| 20 | | | | | | | | | | | |
| 21 | | | | | | | | | | | |
| 22 | | | | | | | | | | | |
| 23 | | | | | | | | | | | |
| 24 | | | | | | | | | | | |
| 25 | | | | | | | | | | | |
| 26 | | | | | | | | | | | |
| 27 | | | | | | | | | | | |
| 28 | | | | | | | | | | | |
| 29 | | | | | | | | | | | |
| 30 | | | | | | | | | | | |
| 31 | | | | | | | | | | | |
| 32 | | | | | | | | | | | |
| 33 | | | | | | | | | | | |
| 34 | | | | | | | | | | | |
| 35 | | | | | | | | | | | |
| 36 | | | | | | | | | | | |

Vista de datos | Vista de variables

IBM SPSS Statistics Processor está listo | Unicode:ON

Escribe aquí para buscar | ESP | 12:28 p. m. 13/08/2020

# CAPÍTULO 2
# ORGANIZACIÓN DE DATOS

## INTRODUCCIÓN

Tiene por objetivo presentar los datos cualitativos o cuantitativos de una muestra, para una o más variables de interés, en tablas de distribución de frecuencias y sus correspondientes gráficos. Con la finalidad de resumir los datos de la muestra, para un mejor análisis e interpretación.

## 2.1. ORGANIZACIÓN DE DATOS

Los datos recopilados de una población, lo cual conforma la muestra al azar, para el estudio de una variable de interés, conducen muchas veces a una gran cantidad de números o atributos, los cuales al intentar analizarlos en su forma original presentan dificultad en cuanto a su análisis e interpretación. Existe muchas interrogantes sobre la naturaleza de los datos, tales como:

a. Que dato o grupos de datos se presenta con la mayor o menor frecuencia.
b. Cual es la distribución empírica de estos datos con respecto a la población (simétrica o asimétrica).
c. La presencia o no de datos discordante (atípicos).
d. Ubicar las medidas de tendencia central, que representan al conjunto de datos.

Estas interrogantes se resuelven si los datos son organizados y presentados en tablas de distribución de frecuencias (univariadas, bivariadas, etc.).

Una tabla de distribución de frecuencias es un arreglo rectangular en filas y columnas en la cual los datos de una muestra o población son resumidos. En la primera fila de la tabla se colocan: el nombre de la variable en estudio, la frecuencia absoluta, la frecuencia relativa, la frecuencia absoluta acumulada, la frecuencia relativa acumulada y la marca de clases, esta última solo para variables cuantitativas continuas.

Titulo:

| Variable en estudio | Frecuencia absoluta | Frecuencia relativa | Frecuencia absoluta acumulada | Frecuencia relativa acumulada | Marca de clase |
|---|---|---|---|---|---|
| Clase-1 | | | | | |
| Clase-2 | | | | | |
| | | | | | |
| | | | | | |
| Clase-k | | | | | |
| Total | | | | | |

Leyenda:
Fuente:

## 2.2. ELEMENTOS DE UNA TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

- TÍTULO

Se coloca en la parte superior central de la tabla, el título debe ser claro y conciso y guarda relación con la variable que se estudia y debe indicar el tamaño de la muestra.

- FUENTE

Se coloca en la parte inferior de la tabla y corresponde al nombre de la entidad (pública o privada) o literatura de donde se obtuvo la información, debe incluir la fecha si es posible. Si el mismo investigador es autor de dicha información, debe colocar fuente propia.

- LEYENDA

Se coloca delante de la fuente y corresponde a una nota o aclaración con respecto al contenido de la tabla. La leyenda se coloca solo cuando es necesario.

- CLASES

Los datos (valores que toma la variable) de la muestra se clasifican en clases. Las clases son cualidades, números o intervalos; si la variable es género, las clases son: varones o mujeres, si la variable es número de hijos por familia las clases son: familias que no tienen hijo 0 hijos, 1 hijo, 2 hijos, etc. Si la variable es edad de estudiantes universitarios las clases podrían

ser 18 hasta 22, 22 hasta 26,..,etc. Las clases para una tabla de distribución de frecuencias deben ser mutuamente excluyentes; es decir, un dato de la muestra solo pude pertenecer a una determinada clase. Al número de clases de una tabla se le representa por la letra "$k$".

Si existe una gran cantidad de datos discretos o continuos, el número de clases debe encontrase entre 5 y 15. Un número pequeño de clases puede ocultar la distribución real del conjunto de datos, mientras que un número muy grande puede dejar sin observaciones algunas de las clases, limitando de esta forma su uso.

- FRECUENCIA ABSOLUTA

Su valor indica el número de veces que los datos de la muestra caen o pertenecen a una determinada clase. Se representa por $f_i$. Para una muestra de tamaño "$n$" y con "$k$" clases se cumple que la sumatoria de todas las frecuencias absolutas es igual a "$n$"; es decir,

$$\sum_{i=1}^{k} f_i = n$$

Siendo

$f_i$ = valor que toma la frecuencia absoluta en la i-esima clase

$n$ = tamaño de la muestra

$k$ = números de clases

- FRECUENCIA RELATIVA

Su valor indica la proporción o porcentaje de la muestra está contenida en una determinada clase (parte de la muestra que pertenece a una determinada clase). Se representa por $h_i$ y se obtiene al dividir la frecuencia absoluta por el tamaño de la muestra; es decir,

$$h_i = \frac{f_i}{n} \quad \text{(En proporción)}$$

$$h_i = \frac{f_i}{n} * 100\% \text{ (En porcentaje)}$$

En ambos casos se cumple que:

Nota: $\sum_{i=1}^{k} h_i = 1 \quad o \quad 100\%$

- FRECUENCIA ABSOLUTA ACUMULADA

Para una variable discreta, un valor de frecuencia absoluta acumulada indica, cuantos datos de la muestra son menores o iguales al valor de la clase correspondiente; para una variable continua, indica cuantos datos de la muestra, tienen valores menores al límite superior para esa clase. Se obtiene al acumular los valores de las frecuencias absolutas. Se representa por $F_i$ , y se obtiene por:

$$F_i = \sum_{j=1}^{i} f_j \qquad i = 1,2,..., k$$

Nota: $F_k = \sum_{j=1}^{k} f_j = n$

$$F_i = f_i + F_{i-1} \qquad F_{-1} = 0$$

- FRECUENCIA RELATIVA ACUMULADA

Para el caso discreto, un valor de frecuencia relativa acumulada, indica la proporción o porcentaje de datos de la muestra que tienen valores menores o iguales al valor de esa clase; para el caso continuo, indica la proporción a porcentaje de datos de la muestra que tienen valores menores al límite superior para esa clase. Se obtiene al acumular los valores de las frecuencias relativas para cada clase. Se representa por $H_i$ .

$$H_i = \sum_{j=1}^{i} h_j \qquad i = 1,2,..., k$$

También se obtiene dividiendo cada valor de la frecuencia absoluta acumulada por el tamaño de la muestra; es decir;

$$H_i = \frac{F_i}{n} \quad i = 1,2,..., k$$

Nota: $H_k = 1 \quad o \quad 100\%$

$$H_i = h_i + H_{i-1} \qquad H_{-1} = 0$$

- MARCA DE CLASE

El concepto de marca de clase, solo se usa cuando se elabora una tabla de distribución de frecuencias para una variable cuantitativa continua, y su interpretación matemática de su valor, corresponde al punto medio para una determinada clase. Estadísticamente, es un valor representativo para todos los datos que pertenecen a una determinada clase, se representa por $x_i'$ Se calcula usando la fórmula:

$$X_i' = \frac{Limite\ \inf erior\ de\ la\ i - esima\ clase + Limite\ \sup erior\ de\ la\ i - esima\ clase}{2}$$

$$X_i' = \frac{LI_i + LS_i}{2}$$

Nota:

En una tabla de distribución de frecuencias con ancho de clase fijo, se cumple que: La segunda, tercera,…k-ésima marca de clase, es igual a la marca de clase anterior más el ancho de clase, es decir:

$$X_{i+1}' = X_i' + A \qquad i = 2,.., k - 1$$

Donde $A =$ ancho del intervalo de clase (ancho de clase)

Demostración

$$X_i' = \frac{LI_i + LS_i}{2} \qquad (1)$$

$$X'_{i+1} = \frac{LI_{i+1} + LS_{i+1}}{2} \qquad (2)$$

$$como \quad LI'_{i+1} = LS_i \quad LS_{i+1} = LS_i + A$$

$$X'_{i+1} = \frac{LS_i + LS_i + A}{2} \qquad (3)$$

$$(3) - (1)$$

$$X'_{i+1} - X'_i = \frac{LS_i + LS_i + A}{2} - \frac{LI_i + LS_i}{2}$$

$$X'_{i+1} - X'_i = \frac{LS_i - LI_i + A}{2} \qquad como \quad LS_i - LI_i = A$$

$$X'_{i+1} - X'_i = \frac{A + A}{2} = A \rightarrow X'_{i+1} = X'_i + A$$

Nota:

$$f_i = F_i - F_{i-1}$$

$$h_i = H_i - H_{i-1}$$

## 2.3. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS PARA UNA VARIABLE CUALITATIVA

Cuando los datos corresponden a una variable cualitativa (Nominal o Jerárquica) la tabla de distribución de frecuencias tiene la siguiente forma:

Título: Corresponde a la variable en estudio

| $i$ | Variable | $f_i$ | $h_i$ |
|---|---|---|---|
| 1 | $C_1$ | $f_1$ | $h_1$ |
| 2 | $C_2$ | $f_2$ | $h_2$ |
| | | | . |
| $k$ | $C_k$ | $f_k$ | $h_k$ |
| | Total | $n$ | 1 |

Fuente: Entidad que proporciona los datos

Los valores $C_i$ corresponden a valores que toma la variable (clase), si la variable es nominal no existe un orden entre sus valores, en cambio si la variable en estudio es ordinal o jerárquica debe tomarse en cuenta el orden de sus valores, ya sea en forma ascendente o descendente.
Para esta variable se presentan algunos tipos de gráficos, pero los más frecuentes es el GRÁFICO de barras verticales u horizontales, GRÁFICO circular, GRÁFICO de líneas y gráficos pictóricos.

Ejemplo
Con la finalidad de estudiar la variable X = Preferencia a determinados marcas de jabones de tocador por las alumnas de la UNMSM, se toma una muestra aleatoria conformada por 30, los resultados son:

| | | | | | |
|---|---|---|---|---|---|
| Rexona | Palmolive | Rexona | Camay | Palmolive | Lux |
| Camay | Lux | Lux | Rexona | Camay | Lux |
| Camay | Rrexona | Camay | Lux | Lux | Camay |
| Lux | Camay | Rexona | Rexona | Camay | Lux |
| Lux | Camay | Palmolive | Lux | Lux | Lux |

Fuente : PACOCHA S . A.

Para este ejemplo la tabla de distribución de frecuencias es:

Preferencias a marcas de jabones de tocador por alumnas de la UNMSM (n=30).

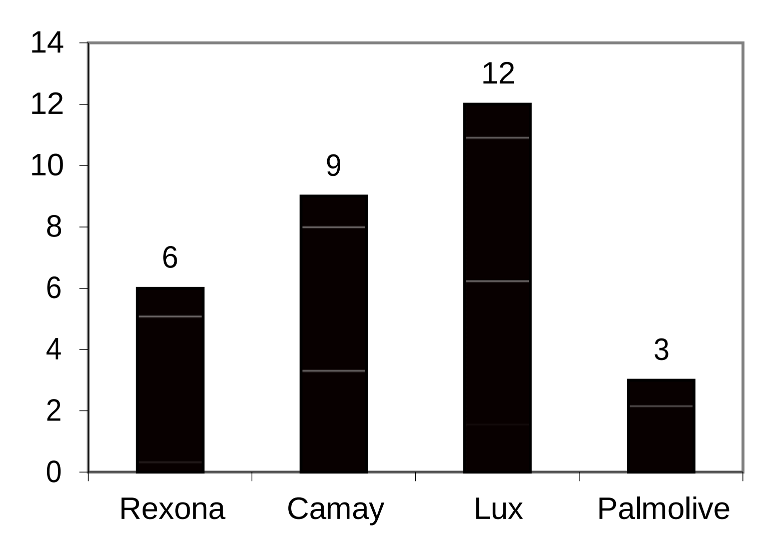
| $i$ | Marca de jabones de tocador | $f_i$ | $h_i$ | $h_i$*100% |
|---|---|---|---|---|
| 1 | Rexona | 6 | 0.2 | 20% |
| 2 | Camay | 9 | 0.3 | 30% |
| 3 | Lux | 12 | 0.4 | 40% |
| 4 | Palmolive | 3 | 0.1 | 10% |
| | Total | 30 | 1.0 | 100% |

Fuente: PACOCHA S.A

Calcular e interpretar:
$f_2 =$ 9 alumnas encuestadas prefieren el jabón de tocador de marca camay
$h_2 =$ 30% de las alumnas encuestadas prefieren el jabón de tocador de marca} Camay.

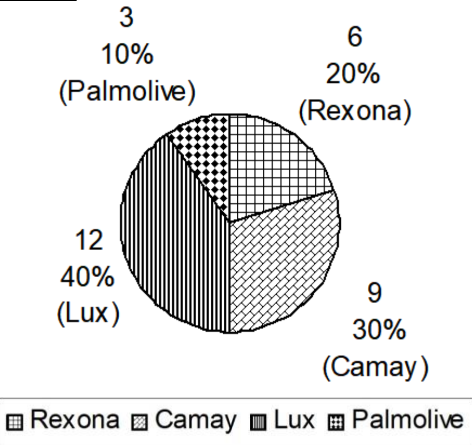
GRÁFICA DE BARRA

Preferencias a marcas de jabones de tocador
Por alumnas de la UNMSM




Fuente: PACOCHA S.A

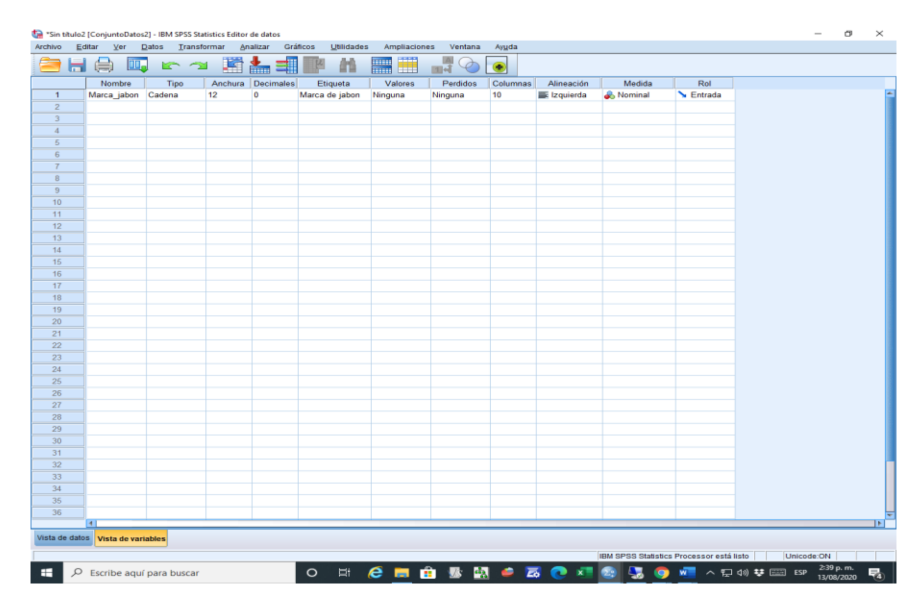
GRÁFICA CIRCULAR



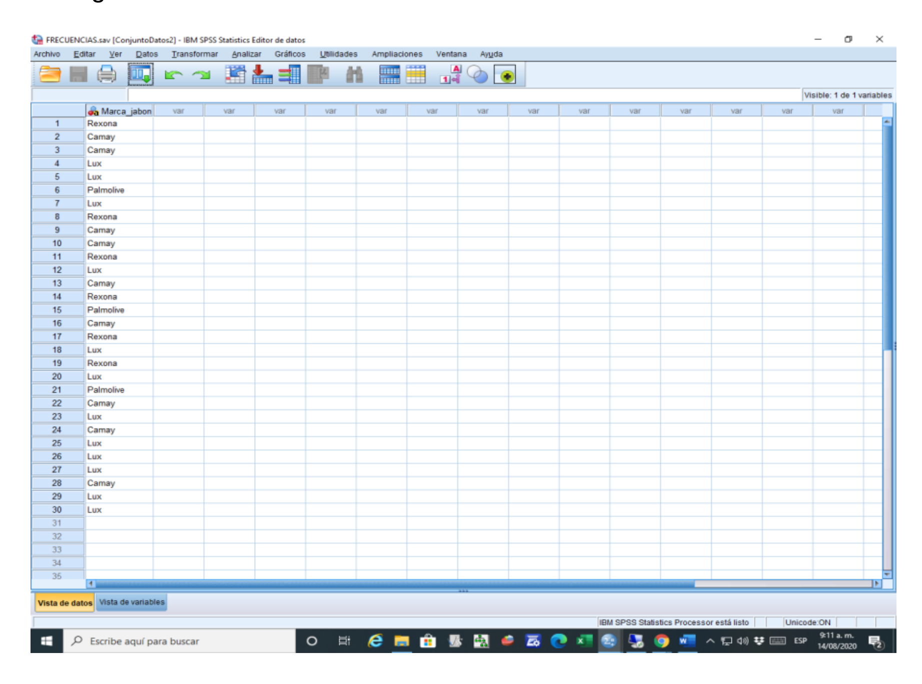

Fuente: Pacocha S.A

Aplicando el SPSS en el Cálculo de las frecuencias, y gráficos.

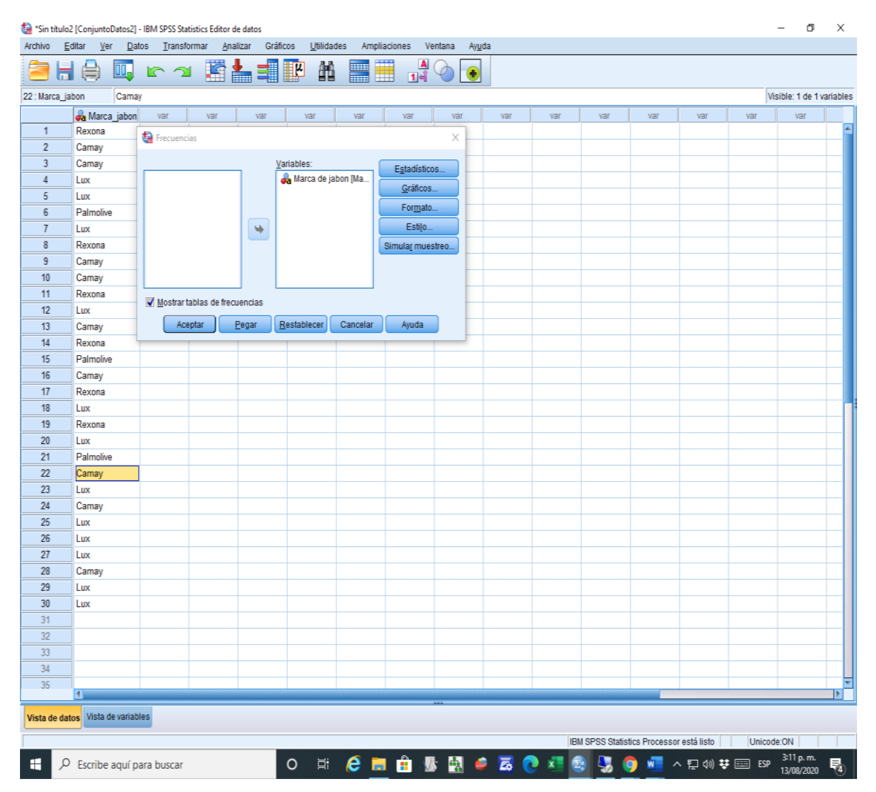
a. Definiendo las variables




b. Ingresando los datos



c. Calculando las frecuencias: En el menú del SPSS, escoger la opción ANALIZAR-ESTADISTICOS DESCRIPTIVOS-FRECUENCIAS-pasar la variable Marca_jabon al casillero de la derecha-y ACEPTAR

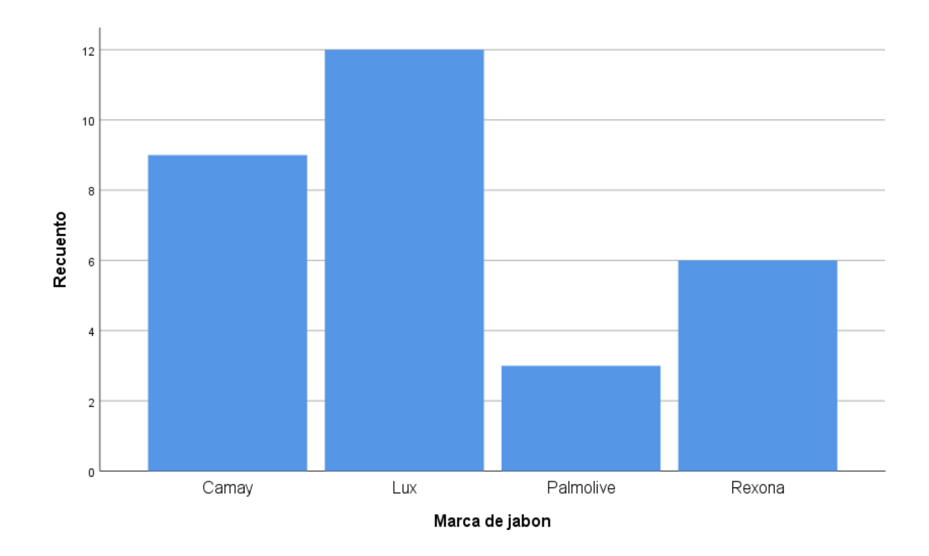



d. Resultados:

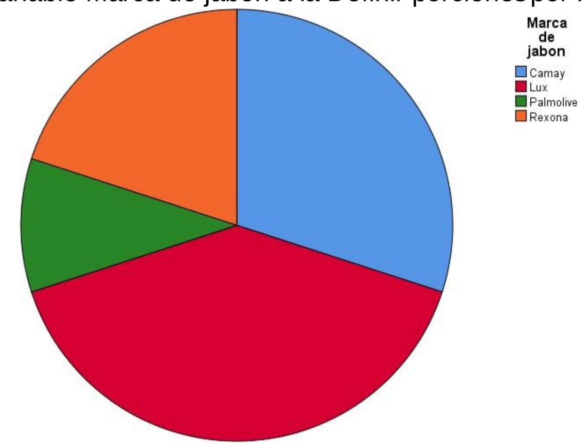
**Marca de jabón**

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---|---|---|---|---|---|
| Válido | Camay | 9 | 30,0 | 30,0 | 30,0 |
| | Lux | 12 | 40,0 | 40,0 | 70,0 |
| | Palmolive | 3 | 10,0 | 10,0 | 80,0 |
| | Rexona | 6 | 20,0 | 20,0 | 100,0 |
| | Total | 30 | 100,0 | 100,0 | |

e. Gráficos barra: En el menú escoger la opción GRÁFICOS-CUADROS DE DIALOGO ANTIGUOS-BAR0RAS-SIMPLES-DEFINIR-pasar la Variable Marca de jabón a la casilla eje de categoría-ACEPTAR




f. Gráficos Circular: Gráficos barra: En el menú escoger la opción GRÁFICOS-CUADROS DE DIALOGO ANTIGUOS-CIRCULAR-RESUMENES PARA GRUPO DE CASOS-DEFINIR-pasar la Variable Marca de jabón a la Definir porciones por-ACEPTAR



## 2.4. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS PARA UNA VARIABLE CUANTITATIVA DISCRETA.

Cuando los datos corresponden a una variable cuantitativa discreta la tabla de distribución de frecuencias tiene la siguiente forma:

Título: corresponde a la variable en estudio

| $i$ | Variable | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|---|
| 1 | $C_1$ | $f_1$ | $h_1$ | $F_1$ | $H_1$ |
| 2 | $C_2$ | $f_2$ | $h_2$ | $F_2$ | $H_2$ |
| .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. |
| $k$ | $C_k$ | $f_k$ | $h_k$ | $F_k$ | $H_k$ |
| | Total | $n$ | 1 | | |

Fuente: entidad que proporciona los datos

Los valores $C_i$ corresponden a valores que toma la variable (clase)

Ejemplo

Los siguientes datos corresponden a una muestra aleatoria de 50 tiendas que se dedican a la venta de gas, con la finalidad de analizar la variable Y = número de balones de gas vendidos en un día determinados, los resultados son:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 21 | 23 | 24 | 23 | 24 | 21 | 24 | 23 | 24 | 22 | 22 | 24 |
| 21 | 22 | 24 | 22 | 24 | 25 | 26 | 25 | 24 | 23 | 23 | 21 |
| 23 | 23 | 22 | 24 | 26 | 26 | 22 | 23 | 22 | 24 | 21 | 25 |
| 22 | 23 | 25 | 26 | 24 | 22 | 25 | 24 | 25 | 21 | 26 | 23 |
| 25 | 24 | | | | | | | | | | |

Elaborar la Tabla de distribución de frecuencias y sus gráficos respectivos

Número de balones de gas vendidos (n=50 tiendas)

| $i$ | Variable | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|---|
| 1 | 21 | 6 | 12 | 6 | 12 |
| 2 | 22 | 9 | 18 | 15 | 30 |
| 3 | 23 | 10 | 20 | 25 | 50 |
| 4 | 24 | 13 | 26 | 38 | 76 |
| 5 | 25 | 7 | 14 | 45 | 90 |
| 6 | 26 | 5 | 10 | 50 | 100 |
| | TOTAL | 50 | 100 | | |

Fuente: SOL GAS S.A

Hallar e interprete:

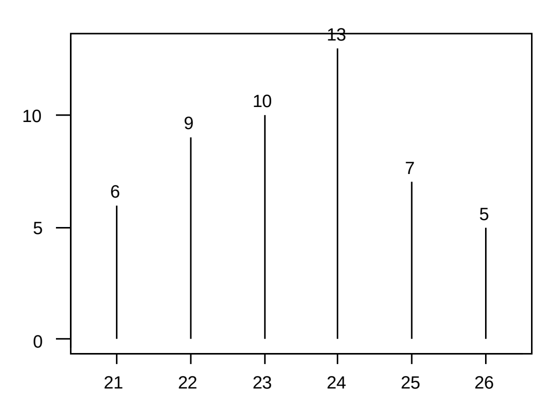
$f_3$ = 10(significa que 10 tiendas, tienen una venta de 23 balones de gas)

$h_2$ =20%( significa que el 20% de las tiendas, tienen una venta de 23 balones de gas)

$F_4$ =38(significa que 38 tiendas, tienen una venta entre 21 y 24 balones de gas o menor o igual a 24 balones de gas)

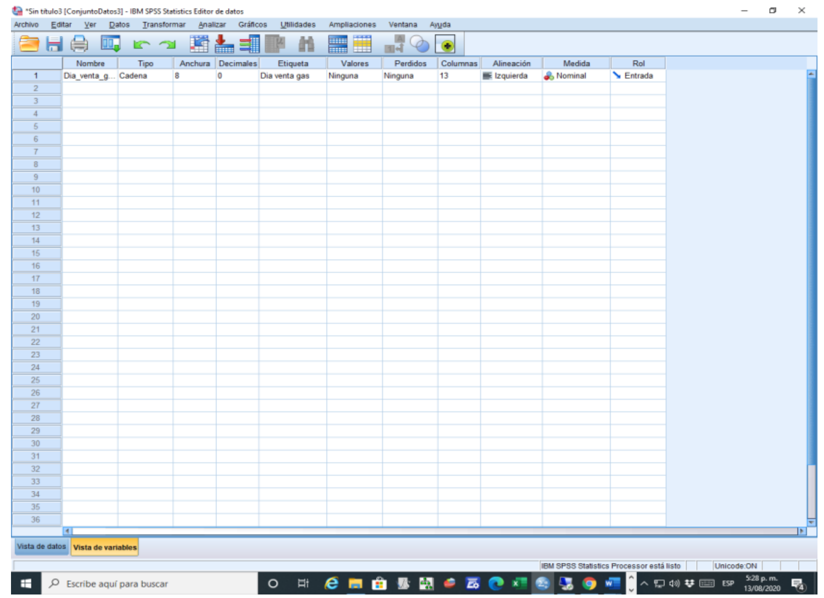
$F_5 - F_2$ =45-15=30(significa que 30 tiendas, tienen una venta entre 23 y 25 balones de gas)

Gráfica de bastones

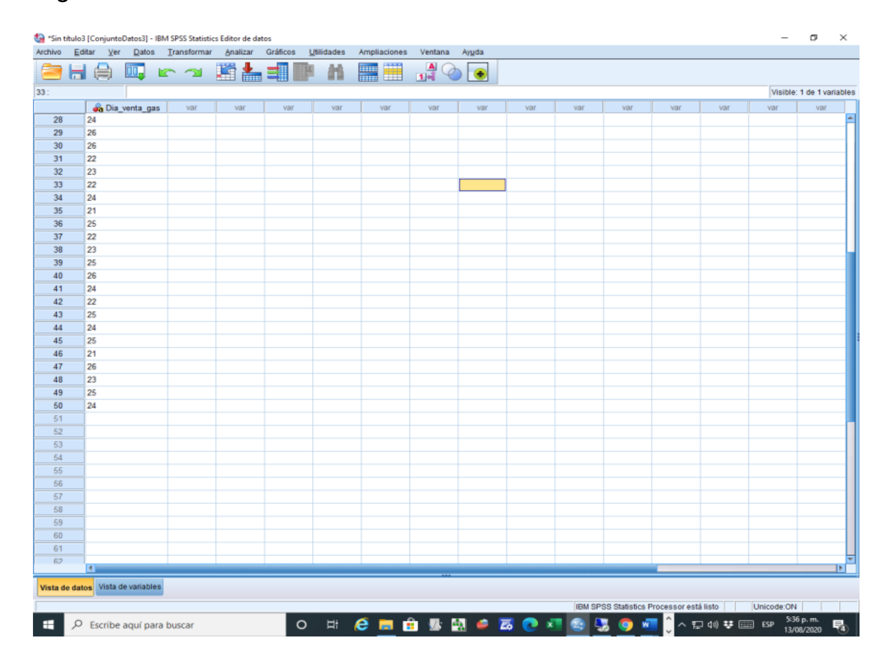



Aplicando el SPSS en el cálculo de las frecuencias, y gráficos.

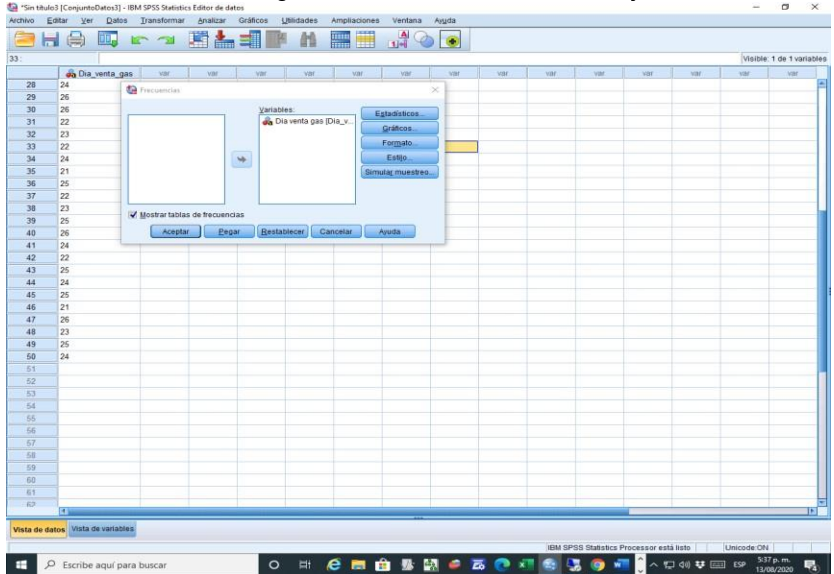
## Definiendo las variables




## Ingresando los datos



Calculando las frecuencias: En el menú del SPSS, escoger la opción ANALIZAR-ESTADISTICOS DESCRIPTIVOS-FRECUENCIAS-pasar la variable días de venta gas al casillero de la derecha-y ACEPTAR

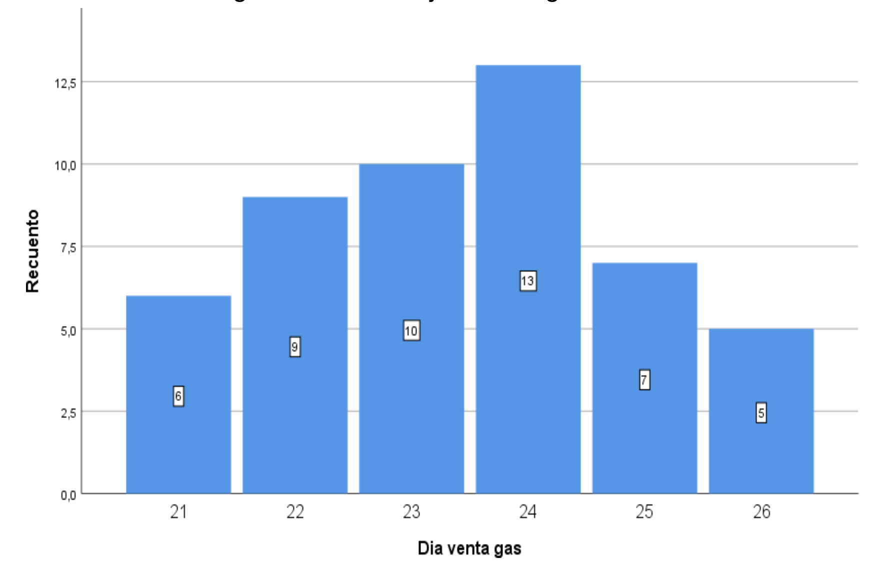



Resultados:

**Dia venta gas**

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---|---|---|---|---|---|
| Válido | 21 | 6 | 12,0 | 12,0 | 12,0 |
| | 22 | 9 | 18,0 | 18,0 | 30,0 |
| | 23 | 10 | 20,0 | 20,0 | 50,0 |
| | 24 | 13 | 26,0 | 26,0 | 76,0 |
| | 25 | 7 | 14,0 | 14,0 | 90,0 |
| | 26 | 5 | 10,0 | 10,0 | 100,0 |
| | Total | 50 | 100,0 | 100,0 | |

Gráficos barra: En el menú escoger la opción GRÁFICOS-CUADROS DE DIALOGO ANTIGUOS-BARAS-SIMPLES-DEFINIR-pasar la Variable Dia venta gas a la casilla eje de categoría-ACEPTAR




## 2.5. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS PARA UNA VARIABLE CUANTITATIVA CONTINUA

Cuando los datos corresponden a una variable cuantitativa continua la tabla de distribución de frecuencias tiene la siguiente forma:

Título: corresponde a la variable en estudio

| $i$ | variable | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|---|
| 1 | $C_1$ | $f_1$ | $h_1$ | $F_1$ | $H_1$ | $X_1'$ |
| 2 | $C_2$ | $f_2$ | $h_2$ | $F_2$ | $H_2$ | $X_2'$ |
| .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | | | |
| K | $C_k$ | $f_k$ | $h_k$ | $n$ | 1 | $X_k'$ |
| | Total | $n$ | 1 | | | |

Fuente: entidad que proporciona los datos

Los valores $C_i$ corresponden a valores que toma la variable (clases)

Elaborar la Tabla de distribución de frecuencia y sus gráficos respectivos

PROCEDIMIENTO PARA ELABORAR UNA TABLA DE DISTRIBUCIÓN FRECUENCIAS PARA UNA VARIABLE CUANTITATIVA CONTINUA

Cuando los datos pertenecen a una variable cuantitativa continua, las clases $C_i$ corresponden a intervalos de la forma $[a\,;\,b)$ , llamados intervalos de clases, siendo "a" y "b" el límite inferior y superior respectivamente del intervalo de clase.

METODOLOGÍA

*1.* Determinación de Rango $R$

$R$ =dato mayor - dato menor

*2.* Determinación de número de clases $K$

$$K = 1 + 3.3\log(n) \quad n = tamaño \quad de \quad la \quad muestra$$

Como el valor de k debe ser un numero entero, usar el redondeo simple

3. Determinación de ancho del intervalo de clase ( $A$)

$$A = \frac{R}{K} \rightarrow \begin{cases} Usar\ redondeo simple,\ con\ el\ mismo \\ número de\ decimales de\ los datos \end{cases}$$

*El redondeosimple casi siempre se comete error* $(E)$

$$E = (A)k - R \begin{cases} i) 0 \rightarrow El\ dato\ de\ Arranque\ para\ la\ tabla = dato\ menor \\ ii) > 0 \rightarrow El\ dato\ de\ arranque\ para\ la\ tabla = dato\ menor - \dfrac{E}{2} \\ iii) < 0 \rightarrow El\ número de\ clases\ se\ incrementa en\ 1\ y\ se\ sigue (i\ ó\ ii) \end{cases}$$

El dato de arranque para la tabla es el primer limite inferior del intervalo de clase.

Ejemplo
Una muestra aleatoria de tamaño n = 34, es tomada con la finalidad de estudiar la variable W = peso del recién nacido en la Maternidad de Lima. Los resultados en Kg. se presentan a continuación:

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.5 | 1.6 | 1.9 | 2.0 | 2.1 | 2.1 | 2.3 | 2.4 | 2.5 | 2.5 | 2.6 | 2.6 | 2.7 | 2.8 | 2.9 | 3.0 | 3.1 |
| 3.1 | 3.2 | 3.2 | 3.2 | 3.2 | 3.3 | 3.3 | 3.3 | 3.4 | 3.4 | 3.5 | 3.6 | 3.6 | 3.7 | 3.7 | 3.9 | 4.1 |

Fuente: Maternidad de Lima.

Metodología

1. Cálculo del Rango” R “.

   R = peso mayor –peso menor.

   R = 4.1-1.5 = 2.6

2. Cálculo del número de clases (K).

$$K = 1 + 3.3Log(34) = 6.05388$$

Como el valor que toma K debe ser un número entero, usar el redondeo simple (redondeo por defecto), se tiene que K = 6.

OBSERVACIÓN

El redondeo simple significa, que si un número tiene como primer decimal un número mayor o igual a 5, entonces la cifra entera del número se incrementa en una unidad.

Ejemplo:

Si, K = 6.7, usando el redondeo simple K = 7;
Si, K = 6.4, usando el redondeo simple K = 6;

a. Cálculo del ancho del intervalo de clase (A)
   El ancho del intervalo de clase se obtiene a través de la siguiente fórmula:

$$A = \frac{R}{K}$$

Para los datos se tiene:

$$A = \frac{R}{K} = \frac{2.6}{6} = 0.43333$$

El valor que toma "A", debe tener el mismo número de decimales de los datos originales. Por redondeo simple se tiene:

$$A = 0.4$$

$$E = AK - R = 0.4(6) - 2.6 = -0.2$$

Como el error es negativo, se debe incrementar en una clase más(k=7)

$$A = \frac{2.6}{7} = 0.4$$

$$E = AK - R = 0.4(7) - 2.6 = 0.2$$

Cuando el error es positivo, la tabla tendrá 7 clases y ancho 0.4

$$valor\ de\ arranque = dato\ menor - \frac{E}{2} = 1.5 - \frac{0.2}{2} = 1.4$$

Para determinar los intervalos de clases es como sigue:

- El valor de 1.4 es el límite inferior de la primera clase, el límite superior de esa clase se obtiene sumándole a 1.4 el ancho de clase 0.4, lo cual da como resultado 1.8.

- El límite inferior del segundo intervalo de clase, corresponde al límite superior del primer intervalo de clase (1.8) y para obtener su límite superior, al límite inferior de ésta clase se le añade del ancho de clase; y así sucesivamente hasta completar las 7 clases.

- Se utiliza un ancho de clase fijo, permite realizar comparaciones de una clase con otra (cuantos elementos de la muestra pertenecen a una determinada clase).

Luego la tabla de distribución de frecuencias para los pesos de los recién nacidos es:

Pesos (kg.) de recién nacidos registrados en la maternidad de Lima (n=34)

| $i$ | $pesos$(kg) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|---|
| 1 | $[1.4-1.8)$ | 2 | 5.88% | 2 | 5.88% | 1.60 |
| 2 | $[1.8-2.2)$ | 4 | 11.76% | 6 | 17.64% | 2.0 |
| 3 | $[2.2-2.6)$ | 4 | 11.76% | 10 | 29.40% | 2.4 |
| 4 | $[2.6-3.0)$ | 5 | 14.71% | 15 | 44.11% | 2.8 |
| 5 | $[3.0-3.4)$ | 10 | 29.42% | 25 | 73.53% | 3.2 |
| 6 | $[3.4-3.8)$ | 7 | 20.59% | 32 | 94.12% | 3.6 |
| 7 | $[3.8-4.2)$ | 2 | 5.88% | 34 | 100% | 4.0 |
| TOTAL | | 34 | 100.0% | | | |

1. Hallar e interpretar los siguientes valores:

$$f_3, h_3, F_3, H_3, H_5 - H_2, X'_2.$$

$f_3$ = 4 (Significa que hay 4 recién nacidos que tienen un peso mayor o igual a 2.2 Kg, pero menor a 2,6 Kg.)

$h_3$ = 11.76%( Significa que el 11.76% de los recién nacidos tienen un peso mayor o igual a 2.2 Kg, pero menor a 2,6 Kg.)

$F_3$ = 10 (Significa que hay 10 recién nacidos que tienen un peso mayor o igual a 1.4 Kg, pero menor a 2.6 Kg.)

$H_3$ = 29.40% (Significa que el 29.40% de los recién nacidos tienen un peso mayor o igual a 1.4 Kg, pero menor a 2.6 Kg.)

$H_5 - H_2$ = 55.89% (significa que el 55.89% de los recién nacidos tienen un peso mayor o igual a 2.2 Kg, pero menor a 3.4 Kg.)

$X'_2$ = 2.0 kg (significa que 2.0 Kg., es el peso representativo para los recién nacidos que tienen un peso mayor o igual a 1.8 Kg., pero menor a 2.2 Kg.) o es un peso representativo para la 2da clase.

2. ¿Cuántos recién nacidos aproximadamente tienen un peso menor a 2.8 Kg.

<u>Por interpolación lineal</u>

2.8 Kg, se encuentra en la cuarta clase; 10 recién nacidos tienen un peso menor a 2.6 kg; 15 recién nacidos tienen un peso menor a 3 kg y “x” recién nacidos tienen un peso menor a 2.8 kg. (utilizando la 2da y 5ta columna)

| limsup. | $F_i$ |
|---|---|
| 2.6 | 10 |
| 2.8 | $x$ |
| 3.0 | 15 |

$$\rightarrow \frac{3.0-2.6}{3.0-2.8}=\frac{15-10}{15-x} \rightarrow x = 12.5 \approx 13$$

Aproximadamente 13 recién nacidos tienen un peso menor a 2.8 kg.

3. ¿Qué porcentaje de los recién nacidos tienen un peso inferior a 2.8 kg.

Como 12.5 recién nacidos tienen un peso inferior a 2.8 Kg, entonces

$$\rightarrow x = 12.5 \quad \rightarrow \frac{12.5}{34} * 100\% = 36.76\%$$

4. ¿Qué porcentaje de los recién nacidos tienen un peso superior a 2.8 kg.

Como 12.5 recién nacidos tienen un peso menor a 2.8 kg, entonces $y$ =34-12.5=21.5 recién nacidos tienen

pesos superiores a 2.8 kg, luego:

$$\rightarrow y = 21.5 \quad \rightarrow \frac{21.5}{34} * 100\% = 63.24\%$$

5. ¿Qué porcentaje de los recién nacidos tienen pesos entre 2.0 kg y 3.6 kg.

$x$

$2.0\ kg \qquad\qquad\qquad 3.6\ kg$

Buscamos cuantos tienen un peso menor a 3.6 kg y cuantos tienen un peso menor a 2.0 kg, la diferencia de ambas cantidades, nos da cuantos recién nacidos tienen pesos entre 2.0kg y 3.6 kg.

Un peso menor a 3.6 kg

| limsup. | $F_i$ |
|---|---|
| 3.4 | 25 |
| 3.6 | $a$ |
| 3.8 | 32 |

$$\rightarrow \frac{3.8-3.4}{3.8-3.6}=\frac{32-25}{32-a} \rightarrow a = 28.5$$

Un peso menor a 2.0 kg

| limsup. | $F_i$ |
|---|---|
| 1.8 | 2 |
| 2.0 | $b$ |
| 2.2 | 6 |

$$\rightarrow \frac{2.2-1.8}{2.2-2.0}=\frac{6-2}{6-b} \rightarrow b = 4$$

$$x = a - b = 28.5 - 4 = 24.5 \rightarrow \frac{24.5}{34}(100\%) = 72.06\%$$

El 72.06% de los recién nacidos aproximadamente tienen pesos entre 2.0 kg y 3.6 kg.

6. ¿Qué peso acumula al 75% de los recién nacidos

| limsup. | $H_i$ |
|---|---|
| 3.4 | 73.53% |
| $x$ | 75.00% |
| 3.8 | 94.12% |

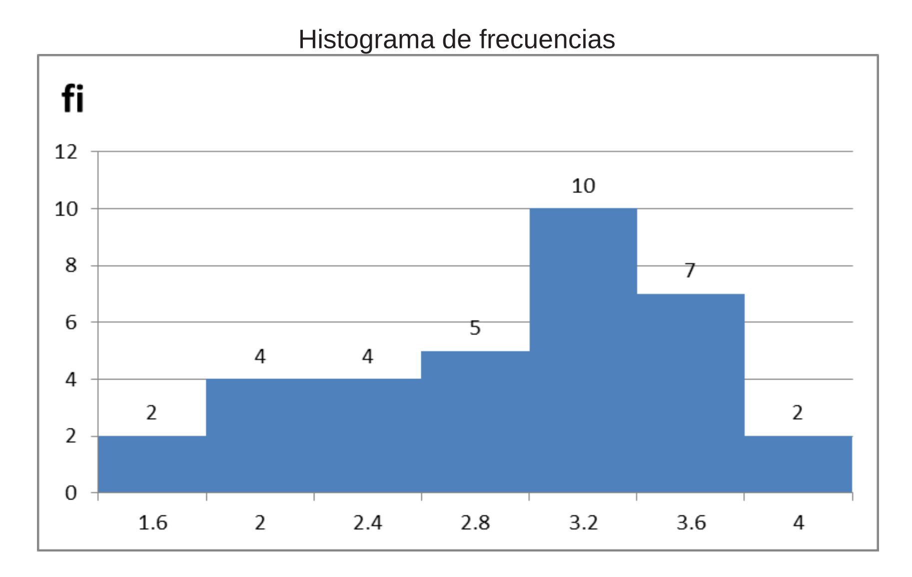
$$\rightarrow \frac{3.8-3.4}{3.8-x}=\frac{94.12\%-73.53\%}{94.12\%-75\%} \rightarrow x = 3.43kg.$$

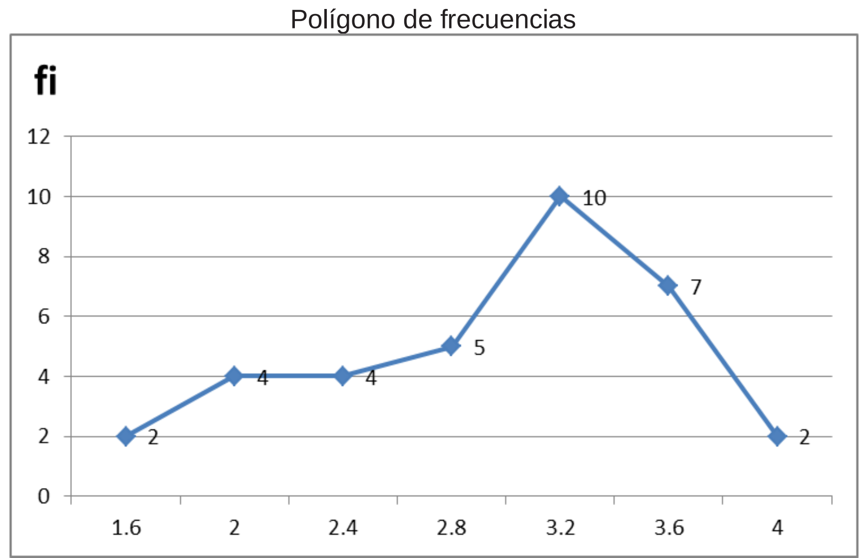
GRÁFICAS:

Histograma de frecuencias

Para construir el histograma de frecuencias, se coloca en el eje de la abscisa los intervalos de clase, y en el eje de la ordenada las frecuencias que pueden ser absolutas o relativas. Al conjunto de rectángulos formados, se le llama histograma de frecuencias que puede ser absoluta o relativa

Para construir el polígono de frecuencias se une los puntos medios de la parte superior de los rectángulos (marcas de clases) y para que este polígono este cerrado se crea marcas de clases ficticias, restándole a la primera marca de clases el valor del ancho de clase y sumándole a la última marca de clases el valor del ancho. Obteniéndose de esta manera un conjunto de líneas quebradas, llamada polígono de frecuencias, que puede ser absoluta o relativa.

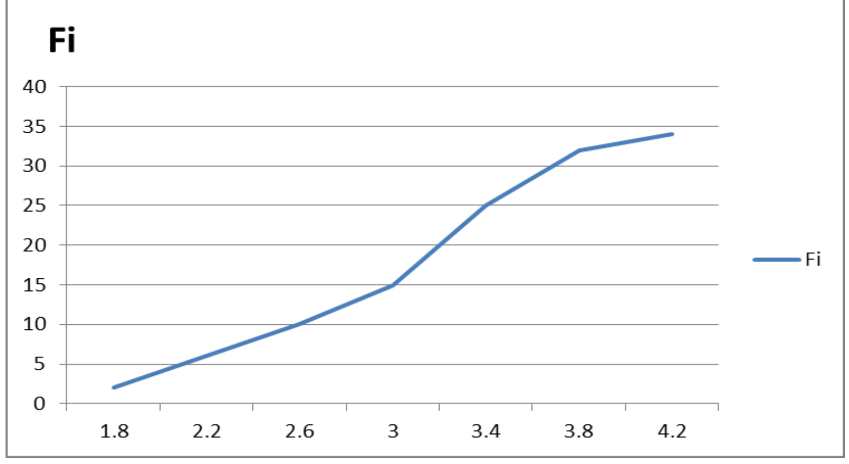



Gráfica de la Ojiva (Menor que)

Para construir la Ojiva en el eje de la abscisa se colocan los intervalos de clases y en eje de la ordenada las frecuencias acumuladas absolutas o frecuencias acumuladas relativas. Por ejemplo la gráfica de la Ojiva para los pesos de los recién nacidos se obtiene de la

siguiente manera, cero recién nacidos tiene un peso menor a 1.4 kg; 2 recién nacidos tienen un peso menor a 1.8 kg; 6 recién nacidos tienen un peso menor a 2.2 kg; 10 recién nacidos tienen un peso menor a 2.6 kg; 15 recién nacidos tienen un peso menor a 3.0 kg; 25 recién nacidos tienen un peso menor a 3.4 kg; 32 recién nacidos tienen un peso menor a 3.8 kg y 34 recién nacidos tienen un peso menor a 4.2 kg, uniendo estos puntos se obtiene la curva creciente a la derecha llamada la Ojiva( “Menor que” ).




Aplicando el SPSS en el cálculo y gráficos.

Definiendo las variables

*Sin título1 [ConjuntoDatos0] - IBM SPSS Statistics Editor de datos

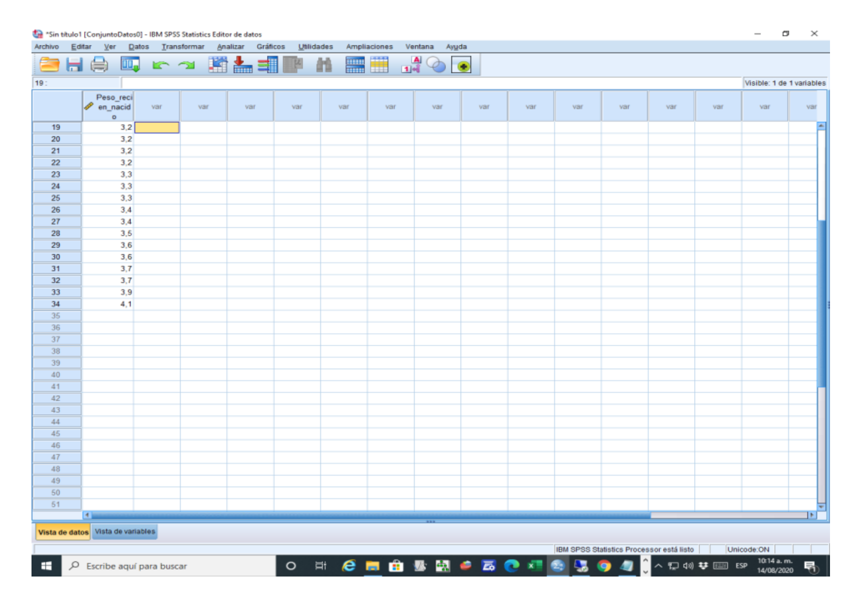
Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

| | Nombre | Tipo | Anchura | Decimales | Etiqueta | Valores | Perdidos | Columnas | Alineación | Medida | Rol |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Peso_recie... | Numérico | 8 | 1 | Peso recien na... | Ninguna | Ninguna | 8 | Derecha | Escala | Entrada |

Vista de datos | Vista de variables

IBM SPSS Statistics Processor está listo | Unicode:ON

Ingresando los datos




Se elabora la tabla de frecuencias para datos agrupados:

1. Cálculo del Rango” R “.

   R = peso mayor –peso menor.

   R = 4.1-1.5 = 2.6

2. Cálculo del número de clases (K).

$$K = 1 + 3.3Log(34) = 6.05388$$

Como el valor que toma K debe ser un número entero, usar el redondeo simple (redondeo por defecto), se tiene que K = 6.

<u>OBSERVACIÓN</u>
El redondeo simple significa, que si un número tiene como primer decimal un número mayor o igual a 5, entonces la cifra entera del número se incrementa en una unidad.

Ejemplo:
Si, K = 6.7, usando el redondeo simple K = 7;
Si, K = 6.4, usando el redondeo simple K = 6;

3. Cálculo del ancho del intervalo de clase (A)

El ancho del intervalo de clase se obtiene a través de la siguiente fórmula:

$$A = \frac{R}{K}$$

Para los datos se tiene:

$$A = \frac{R}{K} = \frac{2.6}{6} = 0.43333$$

El valor que toma "A", debe tener el mismo número de decimales de los datos originales. Por redondeo simple se tiene:

$$A = 0.4$$

$$E = AK - R = 0.4(6) - 2.6 = -0.2$$

Como el error es negativo, se debe incrementar en una clase más(k=7)

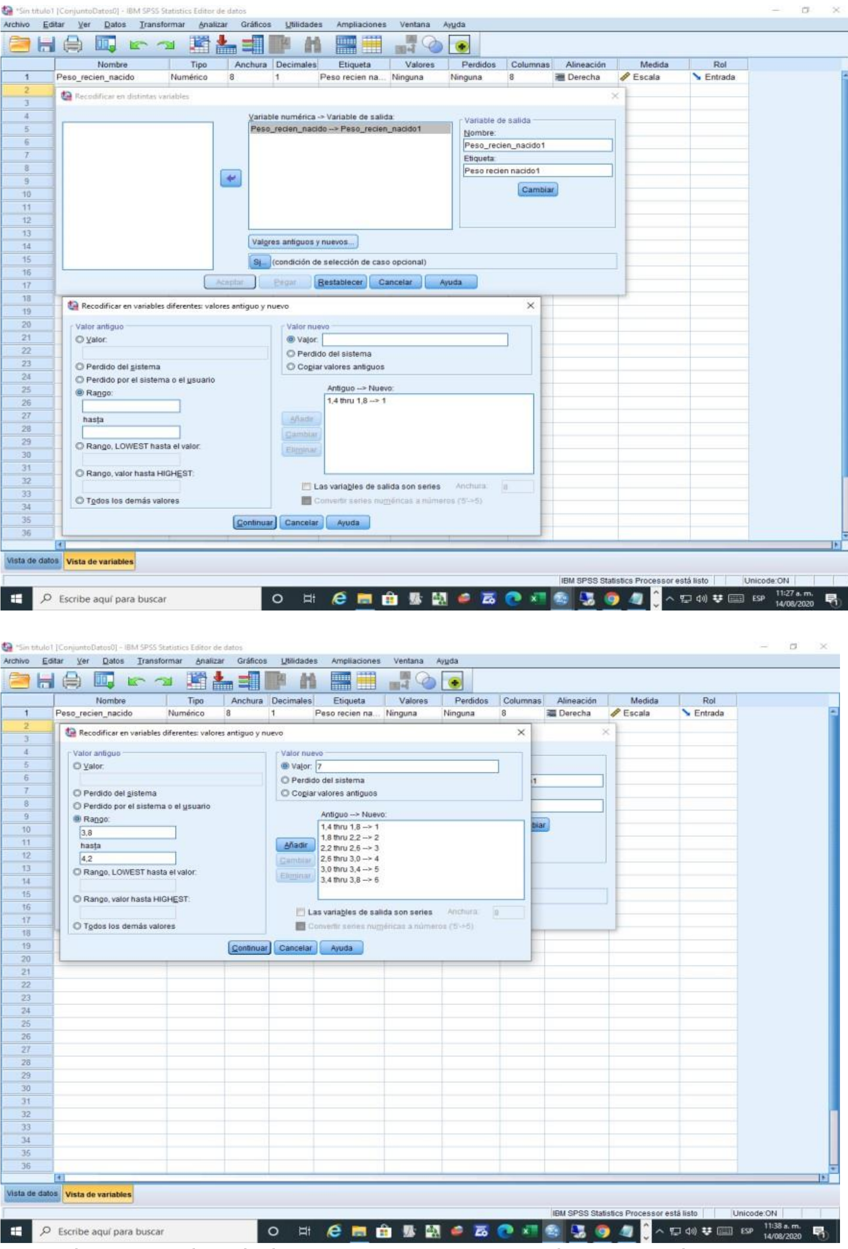
$$A = \frac{2.6}{7} = 0.4$$

$$E = AK - R = 0.4(7) - 2.6 = 0.2$$

Cuando el error es positivo, la tabla tendrá 7 clases y ancho 0.4

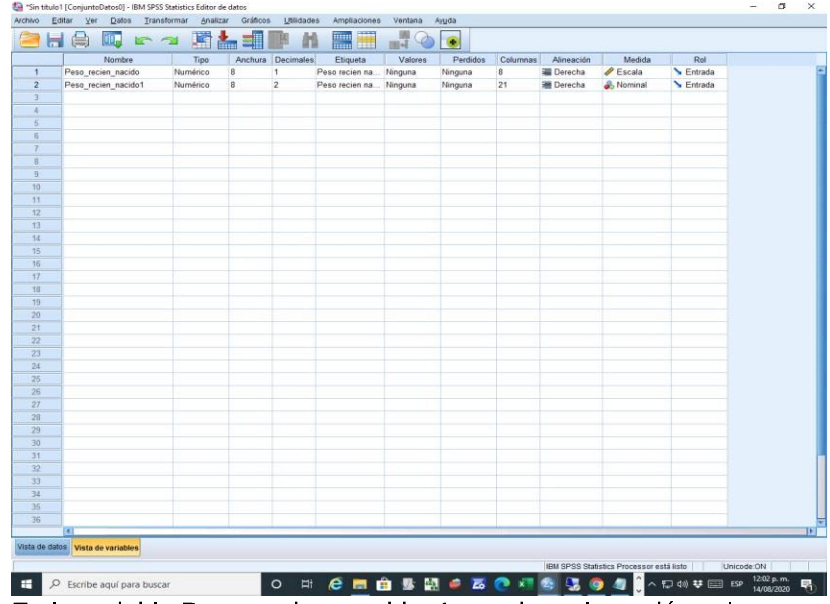
$$valor\ de\ arranque = dato\ menor - \frac{E}{2} = 1.5 - \frac{0.2}{2} = 1.4$$

En el menú del SPSS, escoger la opción TRANSFORMAR-RECODIFICAR EN DISTINTAS VARIABLES-pasar la variable peso-recién-nacido al casillero de la derecha, en la derecha en donde indica variable de salida Nombre: poner peso_recien_nacido1 y en la etiqueta: poner Peso recien nacido1 – PRESIONAL LA OPCION CAMBIAR-VALORES ANTIGUOS Y NUEVOS-RANGO-poner desde 1,4 hasta 1,8, luego en VALOR poner 1 y AÑADIR, así sucesivamente continua con todos los rangos de 3,8 hasta 4,2.

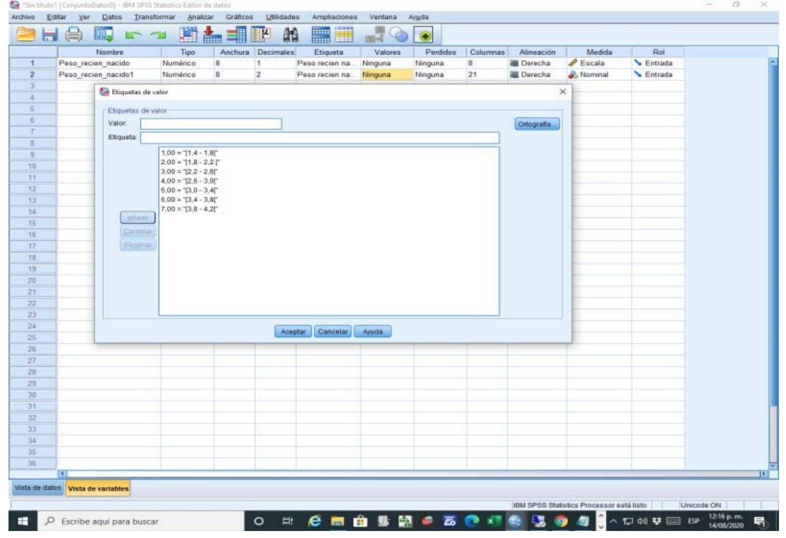



Cuando se termine de ingresar los rangos presionar, continuar y aceptar.

Y se crea una variable Peso_recien_nacido_1.

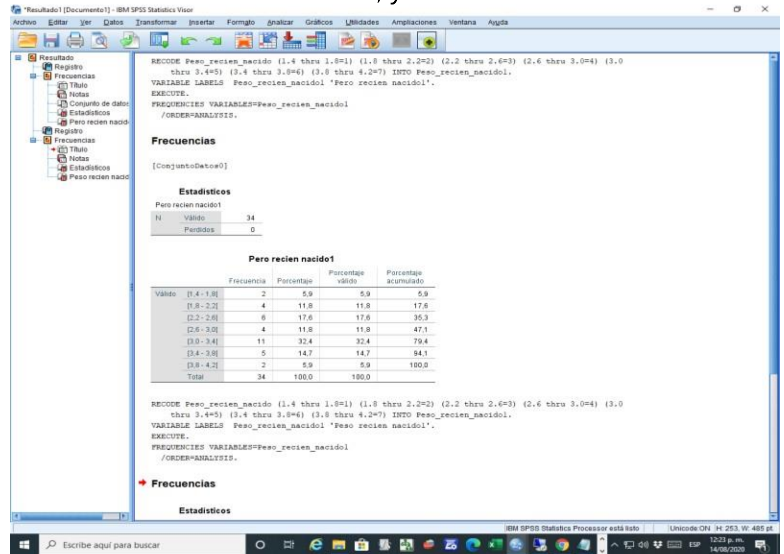


En la variable Peso_recien_nacido_1, presionar la opción valores e ingresar:



Y Poner aceptar.

En el Menú del SPSS escoger la opción ANALIZAR-ESTADISTICOS DESCRIPTIVOS-FRECUENCIA-pasar la variable Peso recien nacido 1 al casillero derecho-ACEPTAR, y nos muestra los resultados



```
RECODE Peso_recien_nacido (1.4 thru 1.8=1) (1.8 thru 2.2=2) (2.2 thru 2.6=3) (2.6 thru 3.0=4) (3.0
    thru 3.4=5) (3.4 thru 3.8=6) (3.8 thru 4.2=7) INTO Peso_recien_nacido1.
VARIABLE LABELS  Peso_recien_nacido1 'Pero recien nacido1'.
EXECUTE.
FREQUENCIES VARIABLES=Peso_recien_nacido1
  /ORDER=ANALYSIS.
```

**Frecuencias**

[ConjuntoDatos0]

**Estadísticos**

Pero recien nacido1

| | | |
|---|---|---|
| N | Válido | 34 |
| | Perdidos | 0 |

**Pero recien nacido1**

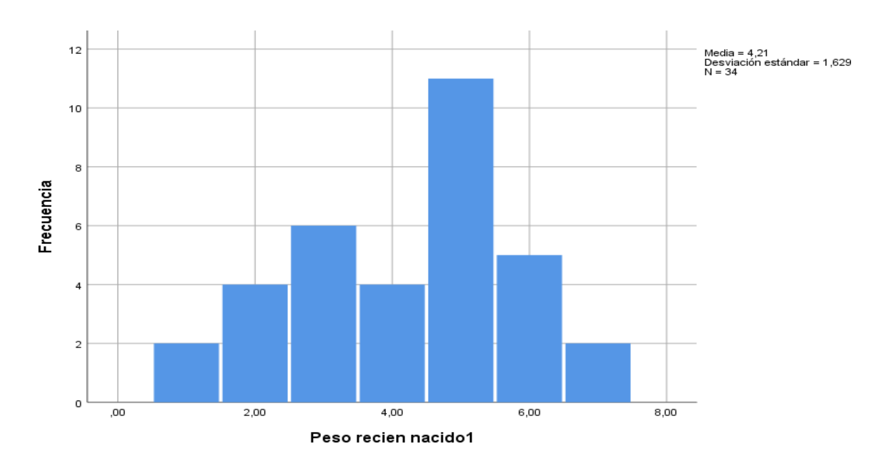
| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---|---|---|---|---|---|
| Válido | [1,4 - 1,8] | 2 | 5,9 | 5,9 | 5,9 |
| | [1,8 - 2,2] | 4 | 11,8 | 11,8 | 17,6 |
| | [2,2 - 2,6] | 6 | 17,6 | 17,6 | 35,3 |
| | [2,6 - 3,0] | 4 | 11,8 | 11,8 | 47,1 |
| | [3,0 - 3,4] | 11 | 32,4 | 32,4 | 79,4 |
| | [3,4 - 3,8] | 5 | 14,7 | 14,7 | 94,1 |
| | [3,8 - 4,2] | 2 | 5,9 | 5,9 | 100,0 |
| | Total | 34 | 100,0 | 100,0 | |

```
RECODE Peso_recien_nacido (1.4 thru 1.8=1) (1.8 thru 2.2=2) (2.2 thru 2.6=3) (2.6 thru 3.0=4) (3.0
    thru 3.4=5) (3.4 thru 3.8=6) (3.8 thru 4.2=7) INTO Peso_recien_nacido1.
VARIABLE LABELS  Peso_recien_nacido1 'Peso recien nacido1'.
EXECUTE.
FREQUENCIES VARIABLES=Peso_recien_nacido1
  /ORDER=ANALYSIS.
```

**Frecuencias**

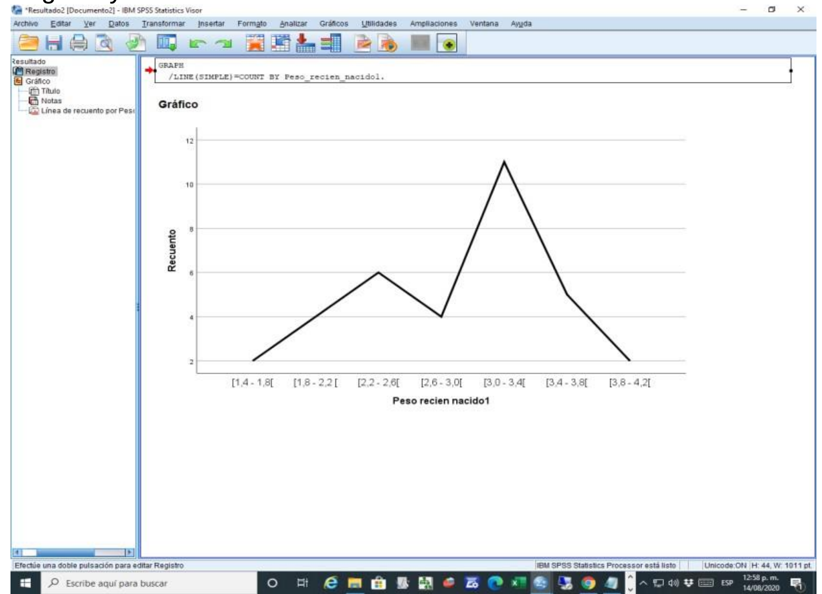
**Estadísticos**

**Histograma de frecuencias:** En el menú del SPSS escoger la opción GRÁFICOS-CUADRO DE DIALOGOS ANTIGUOS-HISTOGRAMA-Pasar la variable Peso recien nacido1 a la casilla variable y ACEPTAR

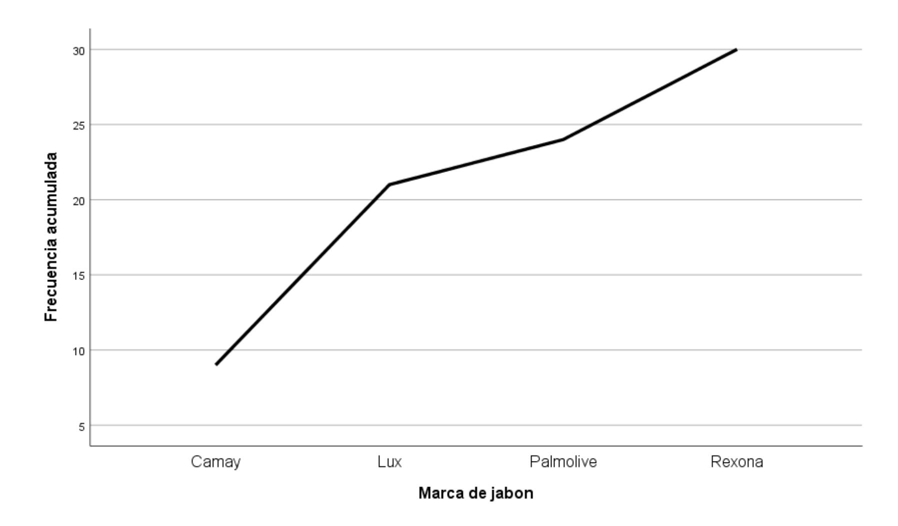



**Polígono de frecuencias**: En el menú del SPSS, escoger la opción: GRÁFICOS-CUADRO DE DIALOGOS ANTIGUOS-LINEAS SIMPLES-

DEFINIR-pasar la variable Peso recien nacido1 a la casilla eje de categoría y marcar N CASOS-ACEPTAR




**Polígono de la Ojiva**: En el menú del SPSS, escoger la opción: GRÁFICOS-CUADRO DE DIALOGOS ANTIGUOS-LINEAS SIMPLES-DEFINIR-pasar la variable Peso recien nacido1 a la casilla eje de categoría y marcar N ACUMULADO-ACEPTAR



Gráfica de la Ojiva ( “Menor que” y “Mayor que”)

Supongamos que la variable en estudios es el tiempo (minutos) que utiliza un usuario en realizar una transacción bancaria

| Tiempo (minutos) | $f_i$ | $h_i$ | $F_i$ | $F_i^*$ | $H_i$ |
|---|---|---|---|---|---|
| 2 - 4 | 3 | 0,10 | 3 | 30 | 0,10 |
| 4 - 6 | 6 | 0,20 | 9 | 27 | 0,30 |
| 6 - 8 | 12 | 0,40 | 21 | 21 | 0,70 |
| 8 - 10 | 7 | 0,23 | 28 | 9 | 0,93 |
| 10 -12 | 2 | 0,07 | 30 | 2 | 1,00 |
| Total | 30 | 1,00 | | | |

$F_i^*$ = valor que se obtiene, al acumular los valores de las frecuencias absolutas empezando por la última clase, así sucesivamente hasta llegar a la primera.
De la tabla, calcular e interpretar:

$F_2 = 9$ , usuarios realizan una transacción bancaria en un tiempo menor a 6 minutos.

$F_4 - F_2 = 28 - 9 = 19$ , usuarios realizan una transacción bancaria, en un tiempo mayor o igual a 6 minutos, pero menor a 10 minutos.

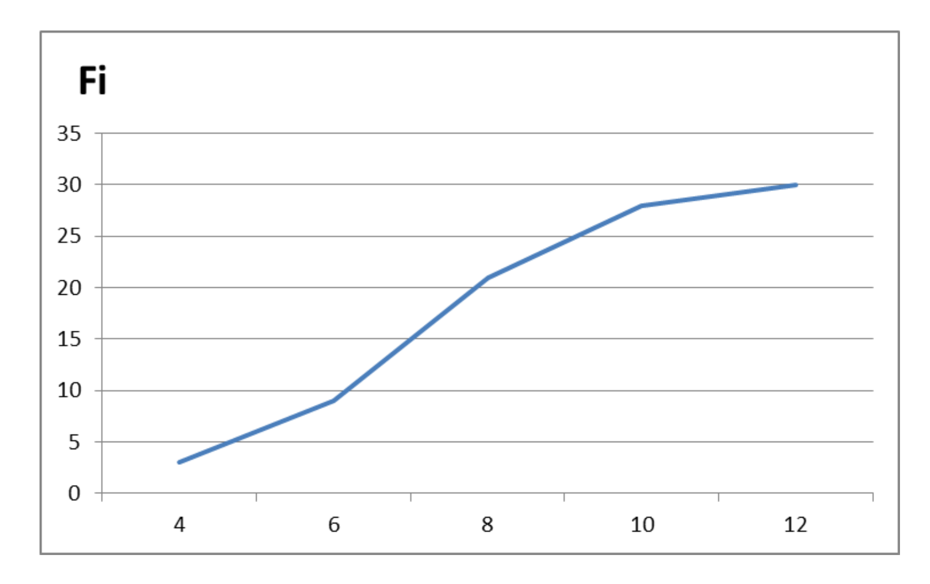
De la tabla anterior, se muestra la ojiva menor que (curva ascendente)

$F_2^*$ =27 usuarios, realizan una transacción bancaria en un tiempo mayor o igual a 4 minutos.

$F_4^*$ =9 usuarios, realizan una transacción bancaria en un tiempo mayor o igual a 8 minutos.

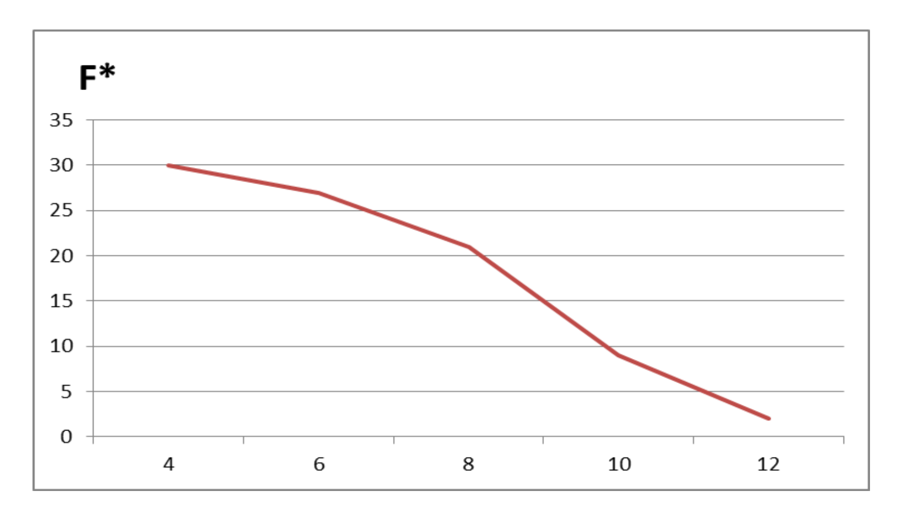
$F_2^* - F_4^*$ =27-9=18 usuarios, realizan una transacción en un tiempo mayor o igual a 4 minutos, pero menor a 8 minutos.

Gráfica de Ojiva menor que




De la tabla anterior, calcular e interpretar:

Gráfica de Ojiva mayor que




Sea la tabla de distribución de frecuencias tiempo que utilizan 50 usuarios en realizar una transacción bancaria.

| Tiempo (minutos) | $f_i$ | $F_i$ | $F_i^*$ |
|---|---|---|---|
| 2 - 4 | 5 | 5 | 50 |
| 4 - 6 | 10 | 15 | 45 |
| 6 - 8 | 20 | 35 | 35 |
| 8 - 10 | 10 | 45 | 15 |
| 10 -12 | 5 | 50 | 5 |
| Total | 50 | | |

Cuantos utilizan un tiempo superior a los 5 minutos

Utilizando la frecuencia absoluta menor que

| $T$ | $F_i$ |
|---|---|
| 4 | 5 |
| 5 | $x$ |
| 6 | 15 |

$$\frac{6-4}{6-5} = \frac{15-5}{15-x} \rightarrow x = 10 \quad usuarios \rightarrow para \quad mayor \quad 50-10=40 \quad Rpta.$$

Utilizando la frecuencia absoluta mayor que

| $T$ | $F_i^*$ |
|---|---|
| 4 | 45 |
| 5 | $x$ |
| 6 | 35 |

$$\frac{6-4}{6-5} = \frac{35-45}{35-x} \rightarrow \frac{2}{1} = \frac{-10}{35-x} \rightarrow x = 40 \quad usuarios \quad Rpta.$$

<u>Ejemplo:</u>
De una muestra de tamaño 50, el dato mayor es 35 y el dato menor 5, determinar las clases para la tabla de distribución de frecuencias.

$$R = 35 - 5 = 30$$

$$K = 1 + 3.3\log(50) = 6.6 \approx 7$$

$$A = \frac{30}{7} = 4.28 = 4$$

$$E = 7x4 - 30 = -2$$

$$A = \frac{30}{8} = 3.75 = 4 \begin{cases} clases \\ 4-8 \\ 8-12 \\ 12-16 \\ 16-20 \\ 20-24 \\ 24-28 \\ 28-32 \\ 32-36 \end{cases}$$

$$E = 8x4 - 30 = 2$$

$$V_a = 5 - \frac{2}{2} = 4$$

NOTA

La siguiente tabla permite determinar el número de clase, para construir la tabla de distribución de frecuencias, utilizando la formula de sturges.

Tabla que proporciona el número de clases
Para un tamaño de muestra dado

| Tamaño "n" de la muestra | Número de clases $K = 1 + 3.3Log(n)$ |
|---|---|
| 20 a 23 | 5 |
| 24 a 46 | 6 |
| 47 a 93 | 7 |
| 94 a 187 | 8 |
| 188 a 376 | 9 |
| 377 a 756 | 10 |
| 757 a 1519 | 11 |
| 1520 a 3053 | 12 |
| 3054 a 6135 | 13 |
| 6136 a 12328 | 14 |
| 12329 a 24770 | 15 |

Fuente: Elaborado por el Autor

Ejemplo aplicativo

La mayor y menor estatura de 50 personas son respectivamente 1.72 mts y 1.55 mts. Calcular el número de clases y su ancho para elaborar la tabla de distribución de frecuencias.

1. Determinación de Rango $R$ =1.72-1.55=0.17
2. Determinación de número de clases

   $K = 1 + 3.3\log(50) = 6.6 = 7$ , usando el redondeo simple

3. Determinación de ancho de clase $A$

$$A = \frac{0.17}{7} = 0.0242 \approx 0.02(redondeosimple)$$

$$Al \quad redondearse\ se\ comete\ error$$

$$E = 0.02 * 7 - 0.17 = -0.03$$

$$Agregamos \quad una \quad nueva \quad clase \quad k = 8$$

$$A = \frac{0.17}{8} = 0.0212 \approx 0.02(redondeosimple)$$

$$E = 0.02 * 8 - 0.17 = -0.01$$

$$Agregamos \quad una \quad nueva \quad clase \quad k = 9$$

$$A = \frac{0.17}{9} = 0.0188 \approx 0.02(redondeosimple)$$

$$E = 0.02 * 9 - 0.17 = 0.01$$

$$dato\ de\ arranque = 1.55 - \frac{0.01}{2} = 1.545 \approx 1.55(redondeosimple)$$

Los intervalos de clases para esta tabla son:

| Estaturas(mts) | $f_i$ |
|---|---|
| 1,55-1.57 | |
| 1,57-1.59 | |
| 1,59-1.61 | |
| 1,61-1.63 | |
| 1,63-1.65 | |
| 1,65-1.67 | |
| 1,67-1.69 | |
| 1,69-1.71 | |
| 1,71-1.73 | |
| Total | 50 |

NOTA-3

No existe una fórmula única para determinar el número de clases (K) en la construcción de una tabla de distribución de frecuencias. Un número muy pequeño de clases tiende a ocultar la distribución real del conjunto de datos, mientras que un número muy grande puede dejar sin observaciones algunas de las clases, limitando de esta forma su uso. La idea es tener un número de clases de tal forma que todos los datos se encuentren en las clases formadas, y además nos permita conocer la distribución empírica de los datos. Se sugiere que el número de ellos debe variar entre 5 y 15.

NOTA-4

Para determinar el número de clases (K) también se pueden usar otras formulas tales como:

$$K = Log_2(n)$$

$$K = \sqrt{n}$$

NOTA-5

Para determinar el número de clases para la tabla, el investigador puede utilizar el criterio anterior.

## 2.6. Ejercicios Propuestos

1. Una serie de 50 datos tiene como variable en estudio, los montos por ventas en miles de nuevos soles, siendo el monto mínimo 85 y el monto mayor 129. Establecer las clases para la tabla.

2. Un conjunto de 50 datos tiene como variable en estudio los pesos de cajas, siendo el peso mínimo 43.43 Kg y el peso Máximo 68.15 Kg, Establecer las clases para la tabla.

3. Los datos que a continuación se presentan, corresponden a una muestra aleatoria de 40 frascos conteniendo mermelada de fresa, el objetivo es estudiar su peso (grs.).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 149 | 164 | 171 | 176 | 181 | 185 | 187 | 190 |
| 156 | 165 | 172 | 176 | 182 | 186 | 187 | 194 |
| 157 | 165 | 174 | 178 | 183 | 186 | 188 | 196 |
| 158 | 168 | 175 | 180 | 183 | 187 | 188 | 198 |
| 162 | 170 | 175 | 180 | 184 | 187 | 189 | 202 |

a. Elaborar la tabla de distribución de frecuencias y hacer los gráficos correspondientes (histograma, polígono y ojiva), use la regla de Sturges

b. Hallar e interprete $H_5 - H_2, \quad 1 - H_4$

c. Usando la tabla de distribución de frecuencias; cual es el peso mínimo que debe tener un frasco, para estar considerado dentro del 15% de los de mayor peso.

4. Los datos que presentan a continuación corresponden a una muestra tomada al azar de 50 personas, con el objetivo de estudiar la variable peso (Kg.):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 38.8 | 39.7 | 39.8 | 42.1 | 42.6 | 42.8 | 42.9 | 43.9 | 44.6 | 45.3 | 45.9 |
| 46.0 | 46.2 | 46.4 | 46.5 | 46.7 | 46.7 | 47.0 | 47.1 | 47.9 | 47.9 | 48.6 |
| 49.0 | 49.5 | 49.5 | 49.5 | 50.3 | 50.8 | 51.3 | 51.3 | 51.3 | 52.2 | 52.3 |
| 52.4 | 52.4 | 52.5 | 53.9 | 54.0 | 54.4 | 54.4 | 54.7 | 54.7 | 55.1 | 55.4 |
| 56.8 | 58.0 | 58.2 | 59.3 | 62.7 | 63.3 | | | | | |

a. Elaborar la tabla de distribución de frecuencias. Usar regla de Sturges.

b. Obtener el histograma de frecuencias absolutas y la ojiva.

c. Calcular e interpretar:

d. $f_3$, $F_4$, $F_5 - F_2$ $X_3'$

e. ¿Cuántos personas aproximadamente tienen un peso menor a 54 Kg?

f. ¿Cuántos personas aproximadamente tienen un peso superior a 52 Kg?

g. ¿Cuántos personas aproximadamente tienen un peso entre 45Kg. y 50 kg?

h. ¿Qué porcentajes de personas tienen un peso inferior a 55 Kg?.

i. ¿Cual es el mínimo peso, para que una persona se encuentre considerado dentro del 30% de los que tienen mayores pesos?

5. Los datos que se presentan a continuación corresponden a los ingresos (nuevos soles) de 45 trabajadores tomados al azar de una empresa.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 695.7 | 721.8 | 728.0 | 735.7 | 752.9 | 758.3 | 761.2 | 780.1 |
| 797.3 | 805.0 | 817.5 | 819.2 | 822.6 | 830.1 | 831.3 | 831.6 |
| 833.6 | 842.0 | 842.1 | 844.9 | 847.1 | 869.3 | 880.7 | 883.9 |
| 889.6 | 895.8 | 896.8 | 897.4 | 899.0 | 919.8 | 920.9 | 921.5 |
| 922.3 | 929.4 | 929.9 | 934.3 | 937.0 | 946.6 | 963.9 | 966.8 |
| 976.0 | 977.8 | 980.2 | 992.2 | 1001.9 | | | |

a. Elaborar la tabla de distribución de frecuencias. Usar la regla de Sturges.

b. Obtener el histograma de frecuencias absolutas y la ojiva.

c. Hallar e interprete.

d. $f_3$ $F_4$ $F_5 - F_2$ $X'_3$

e. ¿Cuántos trabajadores aproximadamente tienen un ingreso menor a S/. 750?

f. ¿Cuántos trabajadores aproximadamente tienen un ingreso superior a S/. 850?

g. ¿Cuántos trabajadores aproximadamente tienen un ingreso entre S/.850 y S/. 950?

h. ¿Qué porcentajes de trabajadores tienen un ingreso inferior a 850 Kg.

i. ¿Cuál es el mínimo ingreso, para que un trabajador este considerado dentro del 30% de los que tienen mayores ingresos?

6. Con la finalidad de realizar un estudio acerca de la cantidad (centenas de Kg) de arroz vendida por semana

en una tienda comercial, se tomó una muestra de 40 semanas, obteniéndose los siguientes resultados.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.47 | 0.79 | 1.36 | 1.46 | 1.51 | 1.65 | 1.67 | 1.82 | 1.95 | 2.05 |
| 2.09 | 2.12 | 2.16 | 2.24 | 2.25 | 2.27 | 2.30 | 2.36 | 2.47 | 2.48 |
| 2.65 | 2.67 | 2.80 | 2.89 | 2.91 | 3.01 | 3.06 | 3.20 | 3.23 | 3.27 |
| 3.36 | 3.51 | 3.56 | 3.60 | 3.69 | 3.72 | 3.89 | 4.15 | 4.47 | 5.15 |

a. Construya la tabla de distribución de frecuencias, utilizar Sturges para determinar el número de clases.
b. Hallar e interprete:

$$f_3 \qquad F_4 \qquad F_5 - F_2 \qquad X'_3$$

c. Calcular e interprete en término del enunciado los valores de la media, mediana y moda.
d. De la tabla de distribución de frecuencias, estime en cuantas semanas de ese año, se vendieron por lo menos 200 Kgs, de arroz pero no mas de 350 Kg.
e. De la tabla de distribución de frecuencias. ¿entre que valores se encuentra el 70% central de la cantidad de arroz vendido?.

7. Se tienen conocimiento que una máquina de tejer está trabajando mal, con la finalidad de conocer el numero de fallas por chompas se extrae una muestra aleatoria de 50 chompas y se contabiliza en numero de fallas, los resultados se presenta a continuación:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 4 | 0 |
| 5 | 2 | 5 | 1 | 4 | 1 | 5 | 3 | 2 | 3 |
| 2 | 2 | 2 | 2 | 2 | 5 | 2 | 5 | 1 | 0 |
| 0 | 0 | 0 | 1 | 0 | 1 | 2 | 5 | 1 | 3 |
| 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 4 | 1 |

a. Elaborar la tabla de distribución de frecuencias
b. Visualizar la tabla a través de gráfico.
c. Hallar e interprete:

$f_3$ $F_4$

8. Los siguientes datos corresponden a la preferencia sobre color de pantalones que usan las alumnas de la Pontifica Universidad Católica del Perú, una muestra aleatoria de tamaño 30 alumnas, da los siguientes resultados:

| | | | | | |
|---|---|---|---|---|---|
| Azul | Rojo | Azul | Azul | Beige | Verde |
| Verde | Verde | Rojo | Blanco | Blanco | Negro |
| Negro | Negro | Negro | Verde | Verde | Verde |
| Verde | Verde | Marrón | Marrón | Marrón | Blanco |
| Rojo | Beige | Azul | Negro | Marrón | Beige |

a. Elaborar la tabla de distribución de frecuencias e interpretar.
b. Realice el GRÁFICO correspondiente.

9. En una encuesta realizada a 50 familias, en la cual la variable en estudio es el número de integrantes de las familias; los resultados son:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 5 | 6 | 3 | 4 | 5 | 6 | 3 | 5 | 4 |
| 3 | 5 | 4 | 3 | 4 | 5 | 3 | 4 | 5 | 2 |
| 3 | 4 | 5 | 3 | 5 | 4 | 3 | 5 | 4 | 3 |
| 3 | 4 | 5 | 4 | 3 | 2 | 6 | 5 | 4 | 4 |
| 2 | 2 | 4 | 5 | 5 | 5 | 5 | 3 | 4 | 6 |

a. Elaborar la tabla de distribución de frecuencias correspondiente
b. Hallar e interprete las clases formadas.

10. Los montos diarios por ventas de una tienda comercial en miles de nuevos soles se presentan a continuación.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1.23 | 1.25 | 1.26 | 1.28 | 2.58 | 2.56 | 2.48 | 2.56 | 2.56 | 2.78 |
| 2.56 | 3.25 | 3.25 | 1.25 | 2.35 | 4.25 | 2.35 | 4.26 | 4.28 | 4.29 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.35 | 4.36 | 2.58 | 1.68 | 1.59 | 1.56 | 1.58 | 1.67 | 2.70 | 2.89 |
| 2.30 | 2.56 | 2.89 | 1.89 | 1.56 | 1.89 | 1.56 | 1.58 | 1.57 | 1.25 |
| 1.23 | 1.23 | 2.58 | 2.47 | 2.31 | 2.56 | 4.56 | 3.25 | 3.24 | 3.25 |
| 3.78 | 3.56 | 3.01 | 3.25 | 3.89 | | | | | |

a. Elaborar la tabla de distribución de frecuencias
b. De la tabla de distribución de frecuencias, cuantos montos por ventas aproximadamente, son superiores a los 2,000 nuevos soles
c. De la tabla de distribución de frecuencias, cuantos montos por ventas aproximadamente se encuentran entre 2,500 nuevos soles y los 4,000 nuevos soles.
d. ¿Que porcentajes de los montos por ventas son superiores a los 3,000 nuevos soles.
e. ¿Cuál es el mínimo monto por venta, para que una venta este considerado dentro del 25% de los mayores montos.

11. Los siguientes datos corresponden a los tiempos en minutos, necesario para que 50 clientes de un banco comercial lleven a cabo una transacción bancaria:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2.3 | 2.4 | 3.3 | 1.8 | 7.8 | 3.1 | 2.4 | 0.4 | 4.2 | 6.3 |
| 0.2 | 4.4 | 9.7 | 4.7 | 0.8 | 3.7 | 4.6 | 1.3 | 1.2 | 7.6 |
| 2.9 | 5.8 | 2.5 | 0.7 | 0.9 | 7.2 | 3.8 | 1.1 | 0.5 | 1.4 |
| 0.4 | 2.8 | 5.6 | 6.2 | 0.4 | 1.6 | 1.5 | 5.5 | 6.8 | 0.5 |
| 2.8 | 3.3 | 9.5 | 1.2 | 1.3 | 1.9 | 2.7 | 3.4 | 5.2 | 1.4. |

a. Construir la tabla de distribución de frecuencias
b. Construir la Ojiva
c. ¿Qué Porcentajes de los tiempos en atención al cliente son superior a los 3 minutos.
d. ¿Que porcentajes de los tiempos de atención al cliente se encuentra entre 2 y 4 minutos.

12. Los siguientes datos corresponden a los montos de ventas, en miles de nuevos soles, de 30 vendedores de una compañía de computadoras.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 45.5 | 32.8 | 48.9 | 51.2 | 89.5 | 70.9 | 85.2 | 70.5 | 65.8 | 60.4 |
| 40.2 | 29.3 | 35.6 | 88.2 | 42.9 | 26.9 | 28.7 | 99.8 | 35.6 | 37.8 |
| 44.2 | 32.3 | 55.2 | 50.6 | 25.4 | 31.7 | 36.8 | 45.2 | 25.1 | 39.7 |

a. Construir la tabla de distribución de frecuencias
b. Construir el histograma de frecuencias y la Ojiva
c. ¿Qué porcentaje de las ventas son superiores a los S/.50,500

13. Los siguientes datos corresponden a los resultados de una encuesta tomada a 50 usuarios sobre los servicios que brinda la biblioteca de cierta universidad.
B= bueno R= regular M= malo

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| B | B | R | B | M | R | R | B | B | R |
| R | B | R | B | M | R | R | B | R | R |
| R | M | R | R | B | R | R | M | R | B |
| B | B | B | R | B | B | B | M | B | B |
| B | M | R | B | M | R | R | B | R | B |

a. Elaborar la tabla de distribución de frecuencias e interpretar.
b. Realice los gráficos correspondientes.

14. La tabla de distribución de frecuencias, corresponde a las edades (años) de usuarios que concurren a un centro de información

| Edades | $F_i$ |
|---|---|
| 18-21 | 3 |
| 21-24 | 9 |
| 24-27 | 21 |
| 27-30 | 39 |
| 30-33 | 59 |
| 33-36 | 69 |
| 36-39 | 76 |
| 39-42 | 80 |
| Total | |

a. ¿Qué porcentajes de las edades de los usuarios son superiores a los 25 años?

b. ¿Entre que edades se encuentran el 80% central de las edades de los usuarios?
c. ¿Cual es la máxima edad para que un usuario se encuentre considerado dentro del 20% de los usuarios que tienen menores edades, y cuántos son aproximadamente?

15. La siguiente tabla de distribución de frecuencias, corresponde a los montos por ventas en miles de nuevos soles, de 50 tiendas comerciales dedicadas a la venta de abarrotes durante un mes. Si los montos por ventas se agrupan en 5 clases de ancho fijo igual a 5000.

| Montos por ventas | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|
| | | | 5 | | 21,000 |
| | | | 17 | | |
| | | | 35 | | |
| | | | 45 | | |
| | | | 50 | | |
| Total | | | | | |

Encuentre:

a. Complete la tabla.
b. ¿Qué cantidad de ventas aproximadamente son inferiores a 30,000 nuevos soles?
c. ¿Cuál es el mínimo monto por venta, para que una venta se encuentre considerado dentro del 25% de las ventas mayores?
d. ¿Entre que valores se encuentran el 80% central de los montos por ventas?

16. La siguiente tabla de distribución de frecuencias corresponden a las edades de 30 usuarios de un centro de información:

| Edades | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| [ – ) | | 0.2 | | | 23 |
| [ – ) | | 0.5 | | | 27 |
| [ – ) | | 0.2 | | | |
| [ – ) | | | | | |
| Total | 30 | | | | |

a. Hallar e interprete: $f_2, \quad h_3, \quad H_4 - H_2$
b. ¿Qué cantidad de usuarios aproximadamente tienen edades inferiores a los 29 años.
c. ¿Qué cantidad de usuarios aproximadamente tienen edades superiores a los 30 años
d. ¿Qué cantidad de usuarios aproximadamente tienen edades entre 23 y 27 años.
e. ¿Qué Cantidad y % de usuarios tienen edades superiores a los 27 años?
f. ¿Cual es la mínima edad, para que un usuario se encuentre considerado dentro del 20% de los que tienen edades mayores?

17. La tabla de distribución de frecuencias corresponde a los tiempos que tarda un pedido realizado por los usuarios que concurren a una biblioteca.

| Tiempo (minutos) | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|
| $[2 - 4)$ | | | 5 | |
| $[4 - 6)$ | | | 16 | |
| $[6 - 8)$ | | | 32 | |
| $[8 - 10)$ | | | 44 | |
| $[10 - 12)$ | | | 50 | |
| Total | | | | |

a. ¿Qué cantidad de usuarios reciben un pedido en menos de 3 minutos
b. ¿Qué cantidad de usuarios aproximadamente reciben un pedido después de los 6 minutos.
c. ¿Qué cantidad de usuarios aproximadamente reciben un pedido entre los 4 y 10 minutos.
d. ¿Que cantidad y porcentaje de usuarios reciben su pedido entre 5 y 9 minutos?
e. ¿Que cantidad y porcentaje de usuarios reciben su pedido antes de los 5 minutos o después de los 9 minutos?

18. La menor edad de los usuarios que concurren a un centro de información es de 18.5 años y la mayor es de 41.3 años, si la cantidad de datos de la muestra es 70, establecer los intervalos de clases.

19. La menor edad de los usuarios que concurren a un centro de información es de 16 años y la mayor es de 50 años, si la muestra es de tamaño 85. Establecer los intervalos de clases.

20. Una muestra de 60 alumnos arroja que la estatura menor es de 1.64 mts, y la mayor 1.72 mts. Establecer los intervalos de clases.

21. Una muestra de 90 alumnos arroja que la estatura menor es de 1.63 mts, y la mayor 1.74 mts. Establecer los intervalos de clases.

22. Una muestra de 70 tiendas comerciales arroja que el menor monto por venta recaudado es de S/.4,300 y el mayor S/.23,

23. Una muestra de 50 alumnos arroja que el peso menor es de 64.5 Kg y el mayor 80.5 kg. Establecer los intervalos de clases.

24. La siguiente tabla de distribución de frecuencias corresponde a las edades de 50 usuarios que concurren a un centro de información:

| Edades | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| $[\quad - \quad)$ | | 0.04 | | | 20 |
| $[\quad - \quad)$ | | 0.16 | | | 24 |
| $[\quad - \quad)$ | | 0.40 | | | 28 |
| $[\quad - \quad)$ | | 0.24 | | | |
| $[\quad - \quad)$ | | 0.10 | | | |
| $[\quad - \quad)$ | | 0.06 | | | |
| Total | | | | | |

a. Calcular e interprete: $f_2, \quad h_4, \quad H_5 - H_2$
b. ¿Qué Cantidad y % de usuarios tienen edades inferiores a los 32 años?
c. ¿Qué Cantidad y % de usuarios tienen edades superiores a los 27 años?
d. ¿Cuál es la mínima edad para que un usuario se encuentre considerado dentro del 20% de los que tienen edades mayores?

25. Los datos de la tabla, tiene como variable en estudio la edad de estudiantes (años)

| Edad(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | | | | 14 |
| | | 0.4 | | 0.56 | 18 |
| | 12 | | | | |
| | 8 | | | | |
| | 2 | | | | |
| Total | | | | | |

a. Completar la tabla de distribución de frecuencias
b. Calcular e interpretar: $h_4 \quad y \quad H_3$

26. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio al número de hijos.

| X=número de hijos | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|
| 1 | | 0.1 | | |
| 2 | | 0.2 | | |
| 3 | 28 | | | |
| 4 | 14 | | | |
| 5 | 7 | | | |
| Total | | | | |

a. Completar la tabla de distribución de frecuencias
b. Calcular e interpretar: $h_3 \quad y \quad H_4$

27. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio los montos por ventas registrados durante 58 días por una tienda comercial.

| X=montos por ventas ( S/.) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | | 4 | | 900 |
| | | | 14 | | 1100 |
| | | | 26 | | |
| | | | 46 | | |
| | | | 54 | | |
| | | | 58 | | |
| Total | | | | | |

a. Completar la tabla de distribución de frecuencias

b. Calcular e interpretar: $f_4, \quad h_2, \quad H_5 - H_2, \quad x_3'$

28. Una tienda comercial registra durante 50 días los montos por ventas; siendo el monto menor S/.320.5 y el mayor es de S/.1236.4. Con esta información establecer las clases para elaborar la tabla de distribución de frecuencias.

29. Sea la tabla de distribución de frecuencias

| Edad de estudiantes (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | 0.10 | | | 22.5 |
| | | 0.18 | | | 27.5 |
| | 10 | | | | 32.5 |
| | 12 | | | | |
| | 10 | | | | |
| | 4 | | | | |
| Total | | | | | |

a. Completar la tabla

b. ¿Qué porcentaje de estudiantes, tienen edades superiores a los 32 años?

c. ¿Qué porcentaje de estudiantes, tienen edades entre los 27 y 42 años?

# CAPÍTULO 3

## INTRODUCCIÓN

En este capítulo se estudia las principales medidas de tendencia central o de posición, las cuales se definen como aquellos valores que representan al conjunto de datos de la variable en estudio para la muestra o población. Se calculan para datos agrupados y no agrupados, así como también sus propiedades.

## 3.1. MEDIDAS DE TENDENCIA CENTRAL O DE POSICIÓN

Son valores estadísticos calculados con los datos de la muestra o de la población y que tienden a ubicarse en el centro de la distribución de los datos. A estas medidas se les considera como valores representativos para el conjunto de datos.

Si estos valores son calculados usando los datos de la muestra, se les llama valores estadísticos, estadígrafos o estimados; en cambio si son obtenidos usando los datos de la población, se les llama parámetros.

Las medidas de tendencia central mas usadas son:

1. La media aritmética o simplemente media.
2. La media ponderada
3. La mediana
4. La moda
5. La media Geométrica
6. La media Armónica

## 3.2. LA MEDIA ARITMÉTICA

Es un valor que tiende a ubicarse en el centro de la distribución de los datos, y es considerado un valor representativo para el conjunto de datos cuando los datos presentan poca variabilidad. Su valor se calcula tanto para la muestra (valor estadístico) como para la población (parámetro). Cuando la media aritmética se calcula utilizando datos que no se encuentran en una tabla de distribución de frecuencias, se dice que la media aritmética se obtuvo para datos sin agrupar, en caso contrario la media aritmética se obtuvo para datos agrupados.

- LA MEDIA ARITMÉTICA PARA LA POBLACIÓN

Sea $X$ una variable en estudio que toma valores $x_1, x_2, x_3, ..., x_N$, la media aritmética para la población se simboliza universalmente con la letra griega $\mu$ ( se lee "mu") y es igual a la sumatoria de todos los datos de la población divididos por el total de ellos.

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

NOTA: El valor de la media aritmética para la población existe, pero casi nunca se conoce, el objetivo es estimar su valor, a través de la media aritmética muestral.

Ejemplo:

Supongamos que se tienen todos los pesos en kg de 3,546 varones adultos de una empresa minera, siendo estos:

65.5 70.5 69.5 70.0 71.0 . . .68.5

Calcular e interpretar la media aritmética poblacional.

$$\mu = \frac{\sum_{i=1}^{3546} x_i}{3546} = \frac{65.5 + 70.5 + 69.5 + 70 + 71 + ... + 68.5}{3546} = 68.2\ kg$$

68.2 Kg. Es el peso promedio representativo para todos los trabajadores de la empresa (parámetro), esto significa que existen datos con pesos inferiores o superiores al valor de la media aritmética.

- LA MEDIA ARITMÉTICA PARA LA MUESTRA

Se X una variable en estudio que toma valores $x_1, x_2, x_3, ..., x_n$, la media aritmética para una muestra se simboliza universalmente con la letra $\bar{x}$ ( se lee x-barra) y su valor es igual

a la sumatoria de todos los datos de la muestra divididos por el total de ellos.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Es la media aritmética para datos sin agrupar. Al valor de la media aritmética, también se le conoce como la media aritmética muestral.

Ejemplo

Los siguientes datos corresponden a los pesos (kg), de 8 personas adultas de sexo varón:

60.5 68.5 58,5 70.5 70.0 68.5 72.0 71.0

Calcular e interpretar la media aritmética.

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{60.5 + 68.5 + 58.5 + 70.5 + 70 + 68.5 + 72 + 71}{6} = 67.4Kg.$$

Interpretación:

67.4 Kg, es el peso promedio de los 8 varones adultos, este valor representa a cada peso de los 8 varones; esto significa que existen pesos inferiores o superiores al valor de la media aritmética.

- PROPIEDADES DE LA MEDIA ARITMÉTICA MUESTRAL

La media aritmética para la muestra, tiene las siguientes propiedades:

1. Sea $x_1, x_2, x_3, \ldots, x_n$, una muestra de tamaño " $n$ " y $\bar{x}$, la media aritmética, definimos a una desviación como la diferencia entre un dato cualquiera $x_i$ y su media aritmética $\bar{x}$; es decir:

$$d_i = x_i - \bar{x}$$

Entonces la sumatoria de todos los valores $d_i$ es igual a cero; esto es:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

2. La suma de los cuadrados de las desviaciones con respecto a la media aritmética, es siempre un valor mínimo; es decir,

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \geq 0$$, es siempre un valor mínimo.

3. Sea $x_1, x_2, x_3, ..., x_n$, una muestra de tamaño "n" y $\bar{X}$, la media aritmética, sea la combinación lineal $y_i = ax_i \pm b$; entonces la media aritmética para variable Y es:

$$\bar{Y} = a\bar{X} \pm b$$

4. Sea una muestra de tamaño " $n$ " , dividida en "k" submuestra de tamaño $n_1, n_2, ... n_k$, y $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$, las medias aritméticas de cada submuestra; entonces la media aritmética de la muestra de tamaño de tamaño $n$, se calcula por:

$$\bar{x}_p = \frac{\sum_{i=1}^{n} n_i \bar{x}_i}{\sum_{i=1}^{n} n_i}$$

a esta media aritmética se le conoce como la media ponderada.

- LA MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

Cuando los datos se encuentran en una tabla de distribución de frecuencias la media aritmética se calcula de la siguiente formula:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i' f_i}{n},$$

donde:

$k =$ Número de clases

$x_i' =$ Marca de clase i-ésima clase( corresponde a la ponderación)

$f_i =$ Frecuencia absoluta de la i-ésima clase

Otra formula que se puede utilizar para calcular la media aritmética para datos agrupados, es en función de las frecuencias relativas:

$$\bar{x} = \sum_{i=1}^{k} x_i' h_i$$

Nota:
Las propiedades de la media aritmética para datos sin agrupar, se cumplen también para datos agrupados.

Ejemplo

Sea la tabla para los datos agrupados de los pesos (kg) de los recién nacido.

| $i$ | $pesos$(kg) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|---|
| 1 | $[1.4 - 1.8)$ | 2 | 5.88% | 2 | 5.88% | 1.60 |
| 2 | $[1.8 - 2.2)$ | 4 | 11.76% | 6 | 17.64% | 2.0 |
| 3 | $[2.2 - 2.6)$ | 4 | 11.76% | 10 | 29.40% | 2.4 |
| 4 | $[2.6 - 3.0)$ | 5 | 14.71% | 15 | 44.11% | 2.8 |
| 5 | $[3.0 - 3.4)$ | 10 | 29.42% | 25 | 73.53% | 3.2 |
| 6 | $[3.4 - 3.8)$ | 7 | 20.59% | 32 | 94.12% | 3.6 |
| 7 | $[3.8 - 4.2)$ | 2 | 5.88% | 34 | 100% | 4.0 |
| TOTAL | | 34 | 100.0% | | | |

Hallar e interprete la media aritmética:

$$\bar{X} = \frac{\sum_{i=1} x_i' f_i}{n} = \frac{1.60(2) + 2.0(4) + 2.4(4) + 2.8(5) + 3.2(10) + 3.6(7) + 4(2)}{34} = 2.94Kg.$$

Interpretación:

2.94 Kg. Es el peso promedio representativo para los 34 recién nacidos de la maternidad de Lima. Significa que en la muestra existen pesos inferiores o superiores al valor de la media aritmética.

- VENTAJAS Y DESVENTAJAS DE LA MEDIA ARITMÉTICA

Ventajas

a. Es fácil de calcular e interpretar su valor
b. Su fórmula se presta a manipulaciones algebraicas
c. La suma de las desviaciones es cero
d. La suma de las desviaciones al cuadrado siempre es mínima

Desventajas:

a. Su valor se encuentra afectados por datos extremos (datos muy altos o datos muy bajo); es decir estos datos tienden a alterar su valor.
b. Deja de ser un valor representativo para la muestra cuando los datos presentan mucha variabilidad o dispersión.
c. No se puede calcular en tabla de distribución de frecuencias con extremos abiertos.
d. Para datos con distribución asimétrica, no debe usarse como medida representativa.

## 3.3. LA MEDIA PONDERADA

Sea X una variable en estudio que toma los valores $x_1, x_2, ..., x_n$ y sea los pesos o ponderaciones asociada a cada valor de la variable $w_1, w_2, ..., w_n$, la media ponderada es definida como:

$$\overline{X}_p = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Si la ponderación es igual para todos, la media ponderada es igual a la media aritmética sin ponderar.

Ejemplo:

En una empresa compuesta por 300 trabajadores, el ingreso mensual es como sigue: 250 de ellos son obreros y ganan $ 850.00, 40 trabajan en la administración y ganan $ 950.00 y el resto son profesionales y ganan $ 1500.00. Hallar e interprete el ingreso promedio para los trabajadores de la empresa.

El enunciado presentado en la tabla:

| $i$ | Trabajadores | Ponderación( $w_i$ ) | sueldo( $x_i$ ) |
|---|---|---|---|
| 1 | Obreros | 250 | 850.00 |
| 2 | Administrativos | 40 | 950.00 |
| 3 | Profesionales | 10 | 1500.00 |
| | Total | 300 | |

$$\overline{X}_p = \frac{\sum_{i=1}^{3} w_i x_i}{\sum_{i=1}^{3} w_i} = \frac{250(850.00) + 40(950.00) + 10(1500)}{250 + 40 + 10} = \$\ 885.00$$

Interpretación:

$.885.00, es el ingreso promedio para los trabajadores de la empresa.

## 3.4. LA MEDIANA

Es una medida de tendencia central, que divide al conjunto de datos ordenados de la muestra, en dos partes iguales; es decir el 50% de los datos tendrán valores menores o iguales al valor de la mediana y el otro 50% de los datos con valores superiores al valor de la mediana.

- CÁLCULO DE LA MEDIANA PARA DATOS SIN AGRUPAR

  Sea X una variable en estudio que toma los valores $x_1$ , $x_2$ ,..., $x_n$ , para calcular el valor de la mediana, se ordenan los datos en magnitud en forma ascendente o

descendente, para calcular la mediana debemos tener en cuenta si el número de datos es par o impar; es decir:

$$n = par \rightarrow me = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} \qquad (1)$$

$$n = impar \rightarrow me = x_{(\frac{n+1}{2})} \qquad (2)$$

Donde:

$x_{(i)}$ = valor que toma la variable X en la posición "i"

El valor de la mediana es útil como medida representativa para el conjunto de datos, cuando estos no tienen un comportamiento homogéneo.

Ejemplo-1

La cantidad de estudiantes que concurren a un centro de información durante la semana es:

230 220 250 228 240 1200

Hallar el valor de la mediana de la concurrencia de estudiantes

Ordenando los datos en forma ascendente:

220 228 230 240 250 1200

Utilizando la fórmula (1), por tener un número par de datos

$$n = 6 \rightarrow me = \frac{x_{(3)} + x_{(4)}}{2} = \frac{230+240}{2} = 235 \quad estudiantes$$

Interpretación:

El 50% de los estudiantes que concurren al centro de información durante los días de la semana es menor o igual a 235, y el otro 50% de los estudiantes es superior a 235.

Ejemplo-2

Sea la serie de datos: 5, 3, 2, 1, 2, 5, 4, 3, Hallar la mediana

Ordenando los datos en forma ascendente:: 1, 2, 2, 3, 3, 4, 5, 5

Como n = 8 (número de datos par), entones:

$$me = \frac{x_4 + x_5}{2} = \frac{3+3}{2} = 3$$

Interpretación

El 50 % de los datos son menores o iguales a 3 y el otro 50 % son superiores a 3.

Ejemplo-3

Hallar e interprete el valor de la mediana para las estaturas (mts) de 7 personas adultas.

1.71 1.69 1.72 1.69 1.68 1.70 1.73

Ordenando las estaturas:

1.68 1.69 1.69 1.70 1.71 1.72 1.73

Como n = 7, entonces: $me = X_{(\frac{7+1}{2})} = X_4 = 1.70$

Interpretación:

El 50 % de las personas adultas tienen estaturas menor o iguales a 1.70 mts, y el otro 50 % de los Cadetes tienen estaturas superior a 1.70 mts.

- CÁLCULO DE LA MEDIANA PARA DATOS AGRUPADOS PARA UNA VARIABLE CUANTITATIVA DISCRETA Y CONTINUA

PARA UNA VARIABLE CUANTITATIVA DISCRETA

Ejemplo

En la siguiente tabla de distribución de frecuencias, la variable en estudio X corresponde al número de balones de gas vendidos diariamente por 50 tiendas. Hallar el valor de la mediana del número de balones de gas vendido diariamente.

| $x_i$ | $f_i$ |
|---|---|
| 22 | 9 |
| 23 | 10 |
| 21 | 6 |
| 25 | 7 |
| 26 | 5 |
| 24 | 13 |
| TOTAL | 50 |

Ordenando (de menor a mayor) los valores que toma la variable X, y obtenemos el valor de las frecuencias absolutas acumuladas.

| $i$ | $x_i$ | $f_i$ | $F_i$ |
|---|---|---|---|
| 1 | 21 | 6 | 6 |
| 2 | 22 | 9 | 15 |
| 3 | 23 | 10 | 25 $(x_{16} \rightarrow x_{25})$ |
| 4 | 24 | 13 | 38 $(x_{26} \rightarrow x_{38})$ |
| 5 | 25 | 7 | 45 |
| 6 | 26 | 5 | 50 |
| TOTAL | | 50 | |

$$n = 50(\,par) \rightarrow me = \frac{x_{25} + x_{26}}{2} = \frac{23 + 24}{2} = 23.5$$

Interpretación:
El 50% de las tiendas tienen una venta diaria menor a igual a 23.5 balones de gas, y el otro 50% de las tiendas tienen una venta diaria mayor a 23.5 balones de gas.

<u>PARA UNA VARIABLE CUANTITATIVA CONTINUA</u>

Cuando los datos se presentan en una tabla de distribución de frecuencias, para una variable continua, el valor de la mediana se calcula haciendo uso de la siguiente formula:

$$me = LI_i + (\frac{\frac{n}{2} - F_{i-1}}{f_i})* A$$

Donde:

$LI_i$ =Limite inferior de la clase que contiene al valor de la mediana

$F_{i-1}$ =Frecuencia absoluta acumulada de la clase anterior a la clase que contiene al valor de la mediana

$f_i$ =Frecuencia absoluta de la clase que contiene al valor de la mediana

$A$ =Ancho del intervalo de clase.

Para ubicar la clase que contiene el valor de la mediana se busca en la columna de la frecuencias absoluta acumulada o frecuencias relativa acumulada ( $F_i$ o $H_i$ , ) el primer valor que sea mayor o igual a la mitad del número de datos o al 50% de los mismos; es decir,

$$F_i \geq \frac{n}{2} \quad o \quad H_i \geq 0.5$$

Otra fórmula que se puede utilizar para calcular la mediana para datos agrupados, en función de las frecuencias relativas:

$$me = LI_i + (\frac{0.5 - H_{i-1}}{h_i})* A$$

Ejemplo-1

Hallar e interprete el valor de la mediana para la tabla de datos agrupados de los recién nacidos.

| $i$ | $pesos$(kg) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|---|
| 1 | $[1.4 - 1.8)$ | 2 | 5.88% | 2 | 5.88% | 1.60 |
| 2 | $[1.8 - 2.2)$ | 4 | 11.76% | 6 | 17.64% | 2.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | $[2.2 - 2.6)$ | 4 | 11.76% | 10 | 29.40% | 2.4 |
| 4 | $[2.6 - 3.0)$ | 5 | 14.71% | 15 | 44.11% | 2.8 |
| 5 | $[3.0 - 3.4)$ | 10 | 29.42% | 25 | 73.53% | 3.2 |
| 6 | $[3.4 - 3.8)$ | 7 | 20.59% | 32 | 94.12% | 3.6 |
| 7 | $[3.8 - 4.2)$ | 2 | 5.88% | 34 | 100% | 4.0 |
| TOTAL | | 34 | 100.0% | | | |

$F_i \geq \frac{n}{2} \rightarrow F_i \geq \frac{34}{2} \rightarrow F_i \geq 17$ , entonces la 5ta clase contiene al valor de la mediana $(i = 5)$

$$me = LI_5 + \left(\frac{\frac{n}{2} - F_{5-1}}{f_5}\right)A = 3 + \left(\frac{17 - 15}{10}\right)(0.4) = 3.08Kg.$$

Interpretación:
El 50% de los recién nacidos tienen pesos menores a iguales a 3.08 Kg y el otro 50% de los recién nacidos tienen pesos superiores a 3.08 Kg.

- VENTAJAS Y DESVENTAJAS DE LA MEDIANA

Ventajas

1. Es fácil de calcular e interpretar su valor
2. Se realizan manipulaciones algebraicas.
3. Su valor no se encuentra afectado por datos extremos.
4. Se puede calcular en tablas de distribución de frecuencias con extremos abiertos

Desventaja
La suma de los cuadrados de las desviaciones con respecto a la mediana no es mínima, en comparación con la media aritmética.

- PROPIEDADES DE LA MEDIANA PARA UN CONJUNTO DE DATOS TRANSFORMADOS

Sea $x_1, x_2, x_3, ..., x_n$ , una muestra de tamaño “"$n$" y $m_{e_x}$ el valor de la mediana, sea la combinación lineal $y_i = ax_i \pm b$ ; entonces el nuevo valor de la mediana para la variable Y es:

$$m_{e_y} = am_{e_x} \pm b$$

Siendo: $a$ y $b$ constantes.

Ejemplo-1
Un conjunto de 20 datos tiene por mediana 24, si al conjunto de datos se le incrementa en un 15% más 3 unidades. Hallar el nuevo valor de la mediana.

Sea el $x_i$, valor del i-esimo datos, por el enunciado del problema el valor del iesimo dato transformado es $y_i = 1.15x_i + 3$ y por la propiedad de la mediana, su nuevo valor es:

$$m_{e_y} = 1.15(24) + 3 = 30.6$$

Ejemplo-2
Sea la tabla de distribución de frecuencias

| Edad | $f_i$ | $F_i$ | $x_i'$ |
|---|---|---|---|
| 18-22 | 4 | 4 | 20 |
| 22-26 | 13 | 17 | 24 |
| 26-30 | 19 | 36 | 28 |
| 30-34 | 12 | 48 | 32 |
| 34-38 | 5 | 53 | 36 |
| Total | 53 | | |

De la tabla, si las edades de las 53 personas se incrementan en un 12% mas 2 años, calcular el nuevo valor de la mediana.

De la tabla calculamos el valor de la mediana.

$$m_{e_x} = 26 + \left(\frac{26.5 - 17}{19}\right) * 4 = 28 \text{ años}$$

Sea el $x_i'$, valor de la i-ésima marca de clase para la tabla, por el enunciado del problema el valor de la i-ésima marca de clase transformada es $y_i' = 1.12x_i' + 2$ y por la propiedad de la mediana, su nuevo valor es:

$$m_{e_y} = 1.12m_{e_x} + 2 = 1.12(28) + 2 = 33.36 \text{ años}$$

## 3.5. LA MODA

Sea $X$ una variable en estudio que toma los valores $x_1$ , $x_2$ ,..., $x_n$ , la moda, es aquel valor de la variable que se presenta con mayor frecuencia, se representa por ( $m_o$ ); es decir,

$$mo = valor\ que\ toma\ la\ \mathrm{var}\ iable\ para\ la\ \max\ ima\ frecuencia\ (f_i)$$

El valor de la moda se interpreta como el valor mas frecuente para el conjunto de datos de una muestra o población.

El valor de la moda puede o no existir. Si un conjunto de datos tiene una sola moda a su distribución se le llama unimodal, si tiene dos modas se le llama bimodal y para más modas multimodal

Ejemplo-1

Hallar la moda para el conjunto de datos

12, 16, 18, 18, 18, 20, 20, 20, 20, 20, 20, 24

El dato que se repite con mayor frecuencia es el valor 20, luego la moda es 20.

Ejemplo-2

Calcular la moda para la serie de datos

1, 1, 1, 3, 3, 4, 3

Los datos que se repiten con igual frecuencia son el 1 y 3, luego la distribución tiene dos modas 1 y 3.

Ejemplo-3

Calcular la moda para el conjunto de datos: 4, 2, 3, 7, 5, 9

Como no existe ningún dato que se repite más de una vez, luego el conjunto de datos no tiene moda.

Ejemplo-4

Una tienda vende durante el día tres tipos de conservas: 120 frascos con fresa, 40 frascos con durazno y 50 frascos con piña. Calcular el tipo de conserva más vendida. En este ejemplo la variable en estudio es tipo de conservas; luego los frascos con fresas registraron la mayor frecuencia en venta

- CÁLCULO DE LA MODA PARA DATOS AGRUPADOS

Cuando los datos se encuentran en una tabla de distribución de frecuencias, la formula para calcular la moda es:

$$mo = LI_i + \left(\frac{d_1}{d_1 + d_2}\right)A$$

Donde:

$LI_i$ = Limite inferior de la clase que contiene al valor de la moda.

$d_1 = f_i - f_{i-1}$ = Diferencia entre la frecuencia absoluta de la clase modal y la frecuencia absoluta de la clase anterior.

$d_2 = f_i - f_{i+1}$ = Diferencia entre la frecuencia absoluta de la clase modal y la frecuencia absoluta de la clase siguiente.

$A$ = Ancho de clase

La clase modal corresponde a la clase de mayor frecuencia absoluta o mayor frecuencia relativa.

Otra fórmula que se puede utilizar para calcular la moda para datos agrupados, es en función de las frecuencias relativas:

$$mo = LI_i + \left(\frac{d_1 / n}{d_1 / n + d_2 / n}\right)A$$

$LI_i$ = Límite inferior de la clase que contiene al valor de la moda.

$\frac{d_1}{n} = h_i - h_{i-1} =$ Diferencia entre la frecuencia relativa de la clase modal y la frecuencia relativa de la clase anterior.

$\frac{d_2}{n} = h_i - h_{i+1} =$ Diferencia entre la frecuencia relativa de la clase modal y la frecuencia relativa de la clase siguiente.

La clase modal corresponde a la clase de mayor frecuencia relativa.

Ejemplo

De la tabla de distribución de frecuencias, en la cual se analiza los pesos de los recién nacidos. Hallar e interprete el valor de la moda.

| $i$ | $pesos$(kg) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|---|
| 1 | $[1.4 - 1.8)$ | 2 | 5.88% | 2 | 5.88% | 1.60 |
| 2 | $[1.8 - 2.2)$ | 4 | 11.76% | 6 | 17.64% | 2.0 |
| 3 | $[2.2 - 2.6)$ | 4 | 11.76% | 10 | 29.40% | 2.4 |
| 4 | $[2.6 - 3.0)$ | 5 | 14.71% | 15 | 44.11% | 2.8 |
| 5 | $[3.0 - 3.4)$ | 10 | 29.42% | 25 | 73.53% | 3.2 |
| 6 | $[3.4 - 3.8)$ | 7 | 20.59% | 32 | 94.12% | 3.6 |
| 7 | $[3.8 - 4.2)$ | 2 | 5.88% | 34 | 100% | 4.0 |
| TOTAL | | 34 | 100.0% | | | |

El valor de la moda se encuentra en la 5ta clase, por presentar la máxima frecuencia absoluta, en este caso la distribución es unimodal para esta muestra, luego (i = 5)

$$d_1 = f_i - f_{i-1} = f_5 - f_4 = 10 - 5 = 5$$

$$d_2 = f_i - f_{i+1} = f_5 - f_6 = 10 - 7 = 3$$

$$mo = LI_5 + \left(\frac{d_1}{d_1 + d_2}\right)A = 3 + \left(\frac{5}{5+3}\right) * 0.4 = 3.25Kg.$$

Interpretación: 3.25 Kg. Es el peso mas frecuente para los recién nacidos.

- PROPIEDADES DE LA MODA PARA DATOS TRANSFORMADOS

Sea $x_1, x_2, x_3, ..., x_n$, una muestra de tamaño "$n$" de datos cuantitativos y $m_{o_x}$ el valor de la moda, sea la combinación lineal $y_i = ax_i \pm b$ ; entonces el nuevo valor de la moda para la variable Y es:

$$m_{o_y} = am_{o_x} \pm b$$

Siendo: $a$ y $b$ constantes.

3.6. LA MEDIA GEOMÉTRICA

Es otra medida de tendencia central, y se utiliza cuando se desea promediar tasas de cambio o proporciones. Para un conjunto de datos $x_1, x_2, x_3, ..., x_n$, $\forall x_i > 0$ la media geométrica se define como la raíz enésima del producto de las observaciones; es decir,

$$\bar{x}_g = \sqrt[n]{(x_1).(x_2).(x_3)...(x_n)}$$

$$\bar{x}_g = Anti\log\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$$

Para datos agrupados la fórmula para calcular la media geométrica es:

$$\bar{x}_g = \sqrt[n]{(x'_1)^{f_1}.(x'_2)^{f_2}.(x'_3)^{f_3}...(x'_k)^{f_k}}$$

$$\bar{x}_g = Anti\log\left(\frac{1}{n}\sum_{i=1}^{k} f_i .\log x'_i\right)$$

Siendo: $x_i'$ = valor de la i-ésima marca de clase

$f_i$ =valor de la i-ésima frecuencia absoluta.

$$n = \sum_{i=1}^{k} f_i$$

<u>Ejemplo-1</u>

Una país tiene una población en el año 1980 de 8 millones de habitantes, en el año 1990 fue de 9.2 millones y en el año 2000 de 10.58 millones. Hallar la población media para esos tres años y el incremento anual.

Cálculo de la población media

$\bar{x}_g = \sqrt[3]{(8)(9.2)(10.58)} = 9.2$ Millones de habitantes es la población media

Cálculo del incremento anual promedio

$$\bar{x}_g = \sqrt{\left(\frac{9.2}{8}\right)\left(\frac{10.58}{9.2}\right)} = \sqrt{\frac{10.58}{8}} = 1.15 \quad \rightarrow 115\%$$

15% es el incremento anual promedio de la población para los 3 años.

<u>Ejemplo-2</u>

Las cantidades de artículos vendidos durante los años 1974 al 1976 por la empresa DURAPLAS, se presenta en la siguiente tabla: Hallar el incremento promedio anual

| Años | Cantidad producida | Variación anual |
|---|---|---|
| 1974 | 200 | |
| 1975 | 400 | 2.0 |
| 1976 | 600 | 1.5 |

La variación anual promedio

$$\bar{x}_g = \sqrt[2]{(2)(1.5)} = 1.7320$$

El incremento anual promedio es 1.732-1=0.732 en % es 73.20%

Ejemplo-3

Hallar el promedio geométrico de las siguientes cantidades

8 40 200 1000

$$\bar{x}_g = \sqrt[4]{(8)(40)(200)(1000)} = 89.4427$$

Ejemplo-3

Calcular el promedio geométrico de las siguientes cantidades:

4, 8, 16, 32

$$\bar{x}_g = \sqrt[4]{(4)(8)(16)(32)} = 11.314$$

## 3.7. LA MEDIA ARMÓNICA

Se utiliza para promediar razones que tienen dimensiones físicas, tales como km/hora, promediar el costo medio de artículos comprados con una cantidad fija de dinero. La media Armónica para n valores $x_i > 0$ , se define como el recíproco de la media aritmética de los recíproco de los datos.

$$\bar{x}_a = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Para datos agrupados

$$x_a = \frac{n}{\frac{f_1}{x_1'} + \frac{f_2}{x_2'} + \dots + \frac{f_k}{x_k'}}$$

Siendo: $x_i'$ = valor de la i-esima marca de clase

$f_i$ = valor de la frecuencia absoluta de la i-esima clase

$$n = \sum_{i=1}^{k} f_i$$

Ejemplo-1

Tres vehículos recorren cierto tramo de una autopista con las siguientes velocidades: 90 km/hora, 110 km/hora y 140 km/hora. Hallar la velocidad media.

$$\bar{x}_a = \frac{3}{\frac{1}{90}+\frac{1}{110}+\frac{1}{140}} = 109.7097 km/hora$$

Ejemplo-2

Una persona concurre a tres mercados para consultar el precio de un artículo determinado llevando consigo la suma de 24 nuevos soles. En el primer mercado pude comprar 4 artículos; en el segundo 5 artículos y en el tercero 6 artículos. Hallar el precio promedio del artículo.

| Cantidad de artículos que puede comprar con los 24 nuevos soles | Precio unitario del articulo |
|---|---|
| 4 | 6.0 |
| 5 | 4.8 |
| 6 | 4.0 |

$$\bar{x}_a = \frac{3}{\frac{1}{6}+\frac{1}{4.8}+\frac{1}{4}} = S/.4.8$$

S/. 4.8 es el precio promedio del artículo.

Ejemplo-3

Un representante del INEI, con la misma cantidad de dinero, compra en diferentes establecimientos cantidades de un mismo artículo, tal como se muestra en la tabla siguiente. Hallar el precio promedio del artículo en los diferentes lugares.

| Dinero | Cantidad de artículo comprado | Precio unitario por artículo |
|---|---|---|
| 20 | 10 | 2 |
| 20 | 8 | 2.5 |
| 20 | 14 | 1.42857 |

$$x_a = \frac{3}{\frac{1}{2} + \frac{1}{2.5} + \frac{1}{1.42857}} = S/.1.875$$

Ejemplo-4

Sean a y b dos números enteros positivo, probar que:

$$\bar{x} \geq \bar{x}_g \geq \bar{x}_a$$

Demostración

$$sean \quad a \quad y \quad b \in R^+$$

Entonces:

$$(\sqrt{a} - \sqrt{b})^2 \geq 0$$

$$a + b - 2\sqrt{ab} \geq 0 \rightarrow a + b \geq 2\sqrt{ab} \rightarrow \frac{a+b}{2} \geq \sqrt{ab} \rightarrow \bar{x} \geq \bar{x}_g \qquad (1)$$

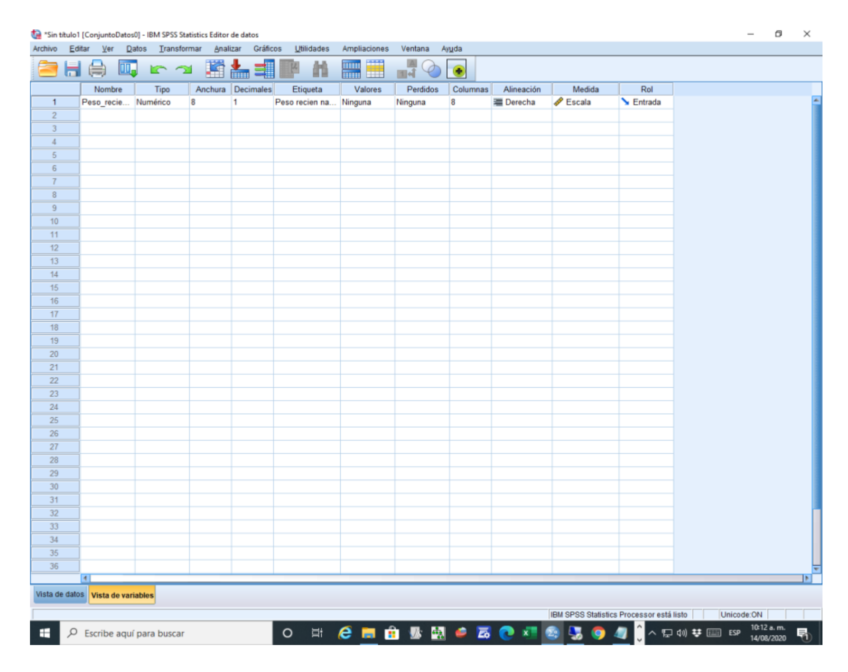
También

$$\left(\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{b}}\right)^2 \geq 0$$

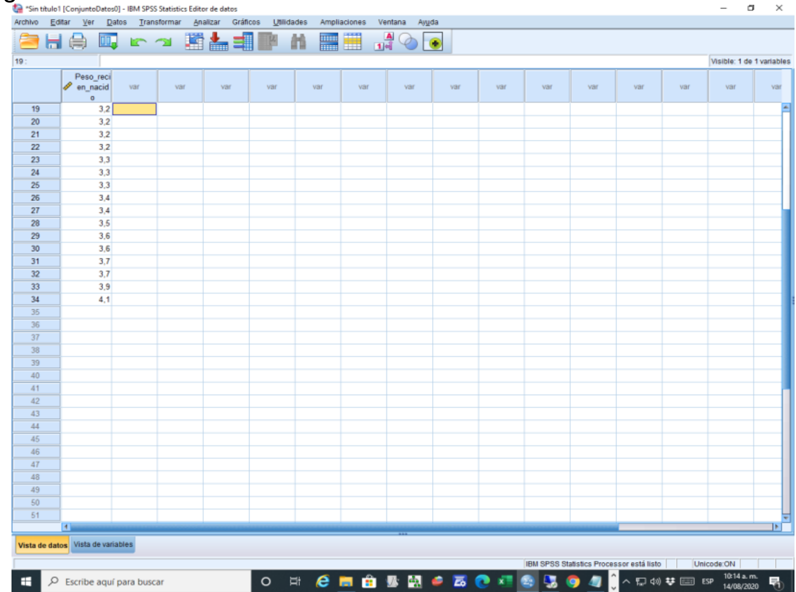
$$\frac{1}{a} + \frac{1}{b} - \frac{2}{\sqrt{ab}} \geq 0 \rightarrow \frac{1}{a} + \frac{1}{b} \geq \frac{2}{\sqrt{ab}} \rightarrow \frac{1}{\frac{1}{a}+\frac{1}{b}} \leq \frac{\sqrt{ab}}{2} \rightarrow \frac{2}{\frac{1}{a}+\frac{1}{b}} \leq \sqrt{ab} \rightarrow \bar{x}_a \leq \bar{x}_g \qquad (2)$$

$$de \quad (1) \quad y \quad (2) \quad \bar{x} \geq \bar{x}_g \geq \bar{x}_a \quad demostrado$$

Aplicando el SPSS para datos no agrupados Media, Mediana, Moda
Definiendo las variables



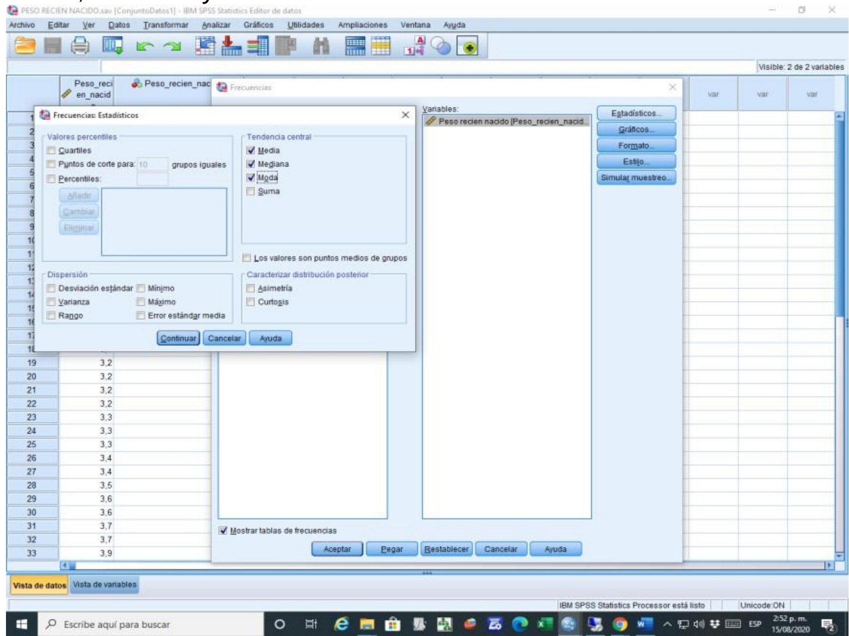

Ingresando los datos




Calculando la Media, La Mediana y la Moda

En el Menú del SPSS escoger ANALIZAR-ESTADISTICOS DESCPRITIVOS-FRECUENCIAS-pasar la variable

Peso_recien_nacido a la casilla de la derecha-ESTADISTICOS-marcar media, mediana y moda-CONTINUAR-ACEPTAR



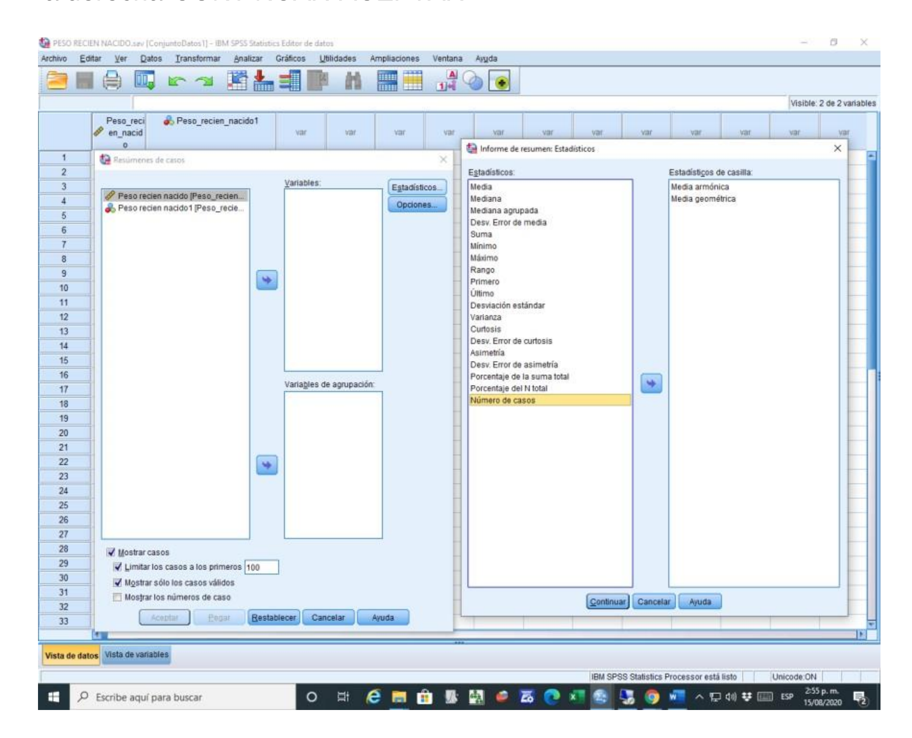

**Estadísticos**

Peso recien nacido

| | | |
|---|---|---|
| N | Válido | 34 |
| | Perdidos | 0 |
| Media | | 2,921 |
| Mediana | | 3,100 |
| Moda | | 3,2 |

Para el cálculo de la Media Geométrica, Media Armónica, con los mismos datos ya ingresados en el menú del SPSS escoger: ANALIZAR-INFORMES-RESUMENES DE CASOS- pasar la variable Peso_recien_nacido a la casilla de la derecha -ESTADISTICOS y

seleccionar media armónica, media geométrica y pasar a la casilla de la derecha-CONTINUAR-ACEPTAR




**Resúmenes de casos**

| Peso recien nacido | |
|---|---|
| Media armónica | Media geométrica |
| 2,748 | 2,839 |

Para datos agrupados
Se elabora la tabla de frecuencias para datos agrupados:

4. Cálculo del Rango” R “.

   R = peso mayor –peso menor.

   R = 4.1-1.5 = 2.6

5. Cálculo del número de clases (K).

$$K = 1 + 3.3Log(34) = 6.05388$$

Como el valor que toma K debe ser un número entero, usar el redondeo simple (redondeo por defecto), se tiene que K = 6.

OBSERVACIÓN

El redondeo simple significa, que si un número tiene como primer decimal un número mayor o igual a 5, entonces la cifra entera del número se incrementa en una unidad.

Ejemplo:

Si, K = 6.7, usando el redondeo simple K = 7;
Si, K = 6.4, usando el redondeo simple K = 6;

6. Cálculo del ancho del intervalo de clase (A)

El ancho del intervalo de clase se obtiene a través de la siguiente fórmula:

$$A = \frac{R}{K}$$

Para los datos se tiene:

$$A = \frac{R}{K} = \frac{2.6}{6} = 0.43333$$

El valor que toma "A", debe tener el mismo número de decimales de los datos originales. Por redondeo simple se tiene:

$$A = 0.4$$

$$E = AK - R = 0.4(6) - 2.6 = -0.2$$

Como el error es negativo, se debe incrementar en una clase más(k=7)

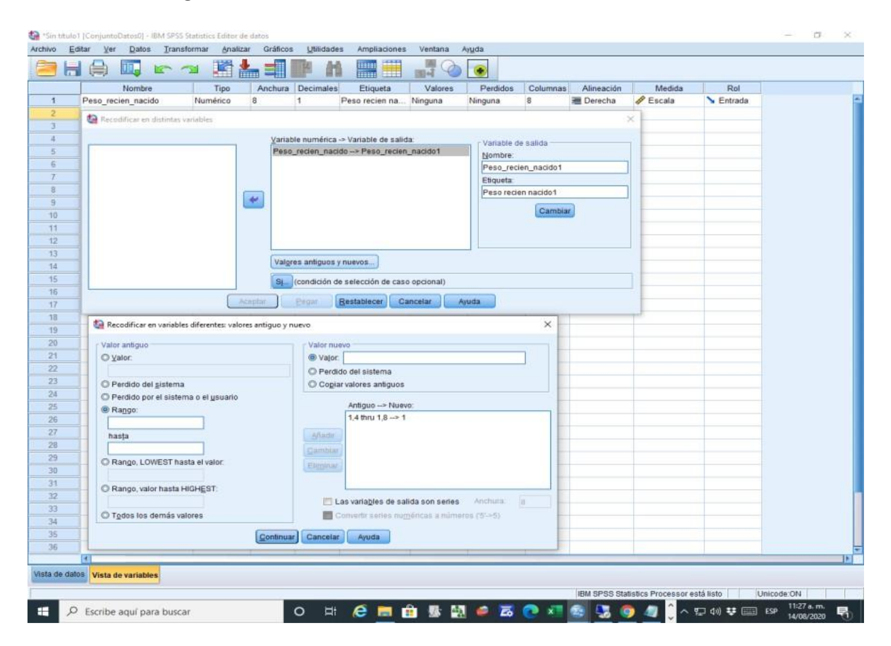
$$A = \frac{2.6}{7} = 0.4$$

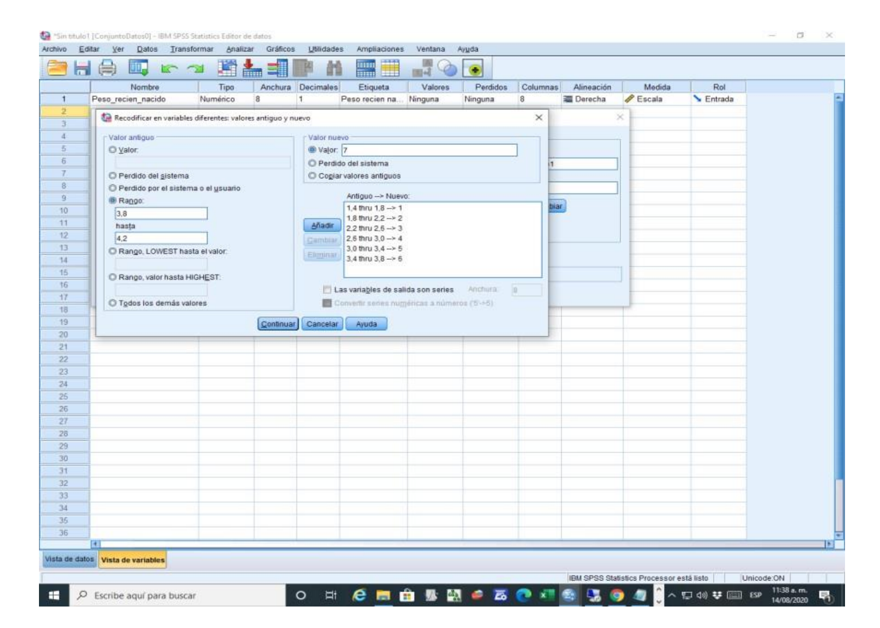
$$E = AK - R = 0.4(7) - 2.6 = 0.2$$

Cuando el error es positivo, la tabla tendrá 7 clases y ancho 0.4

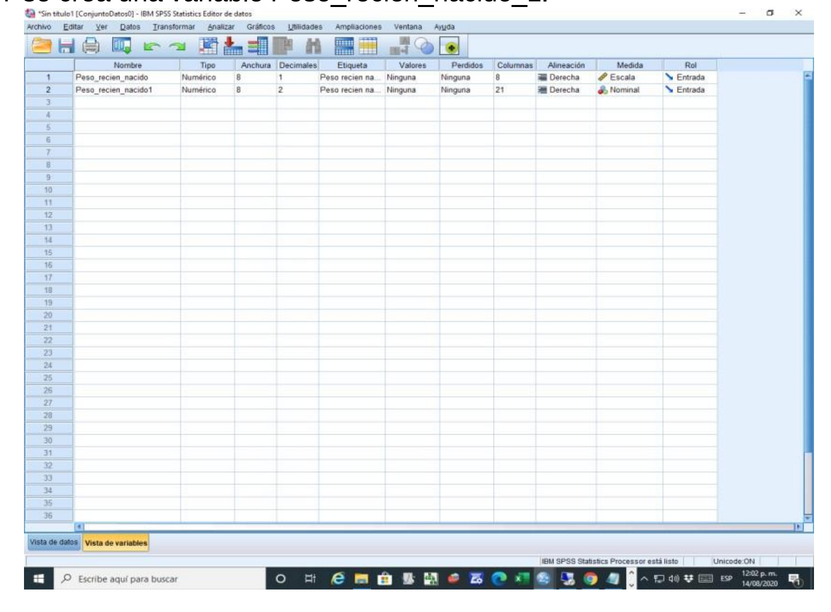
$$\text{valor de arranque} = \text{dato menor} - \frac{E}{2} = 1.5 - \frac{0.2}{2} = 1.4$$

En el menú del SPSS, escoger la opción TRANSFORMAR-RECODIFICAR EN DISTINTAS VARIABLES-pasar la variable peso-recién-nacido al casillero de la derecha, en la derecha en donde indica variable de salida Nombre: poner peso_recien_nacido1 y en la etiqueta: poner Peso_recien_nacido1 – PRESIONAL LA OPCION CAMBIAR-VALORES ANTIGUOS Y NUEVOS-RANGO-poner desde 1,4 hasta 1,8, luego en VALOR poner 1 y AÑADIR, así sucesivamente continua con todos los rangos de 3,8 hasta 4,2.

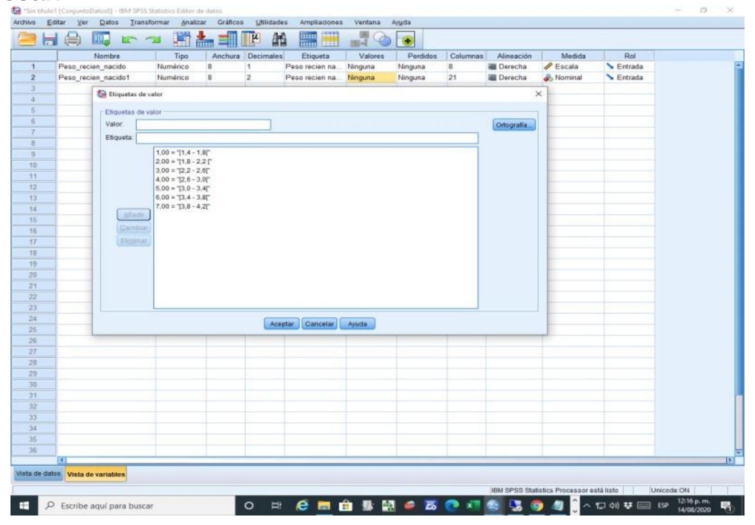





Cuando se termine de ingresar los rangos presionar, continuar y aceptar.
Y se crea una variable Peso_recien_nacido_1.

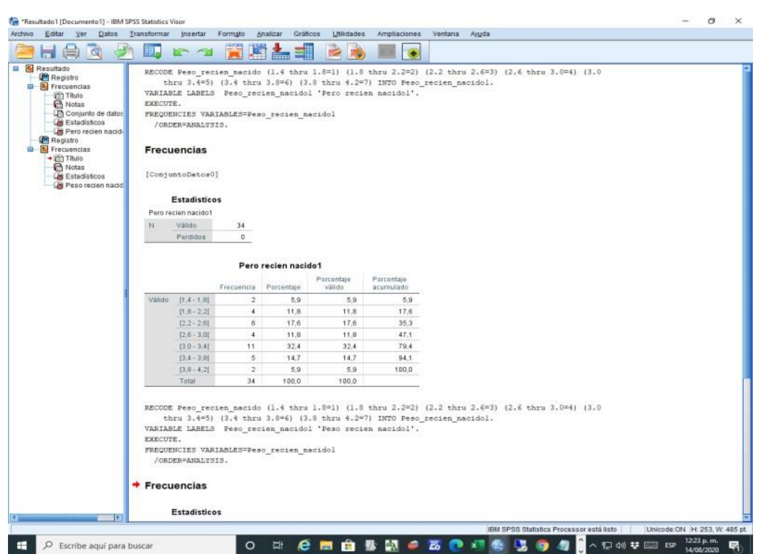


En la variable Peso_recien_nacido_1, presionar la opción valores e ingresar:



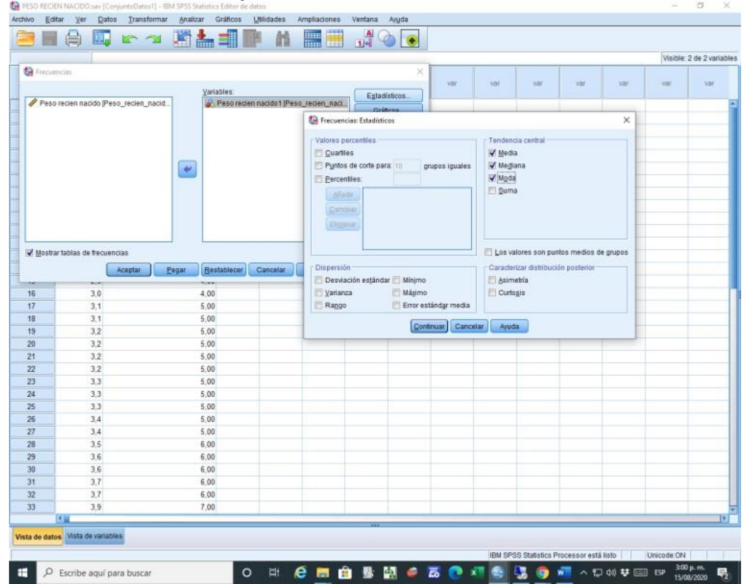

Y Poner aceptar.

En el Menú del SPSS escoger la opción ANALIZAR-ESTADISTICOS DESCRIPTIVOS-FRECUENCIA-pasar la variable Peso_recien_nacido 1 al casillero derecho-ACEPTAR, y nos muestra los resultados




Calculando la Media, La Mediana y la Moda

En el Menú del SPSS escoger ANALIZAR-ESTADISTICOS DESCPRITIVOS-FRECUENCIAS-pasar la variable Peso_recien_nacido1 a la casilla de la derecha-ESTADISTICOS-marcar media, mediana y moda-CONTINUAR-ACEPTAR

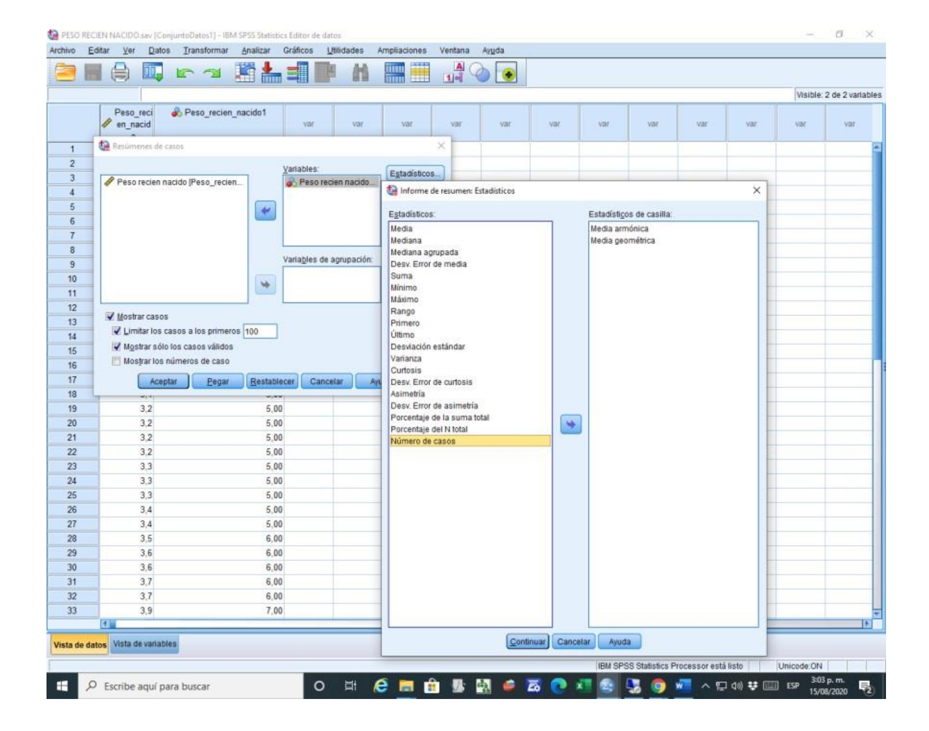



**Estadísticos**

Peso recien nacido1

| | | |
|---|---|---|
| N | Válido | 34 |
| | Perdidos | 0 |
| Media | | 4,2059 |
| Mediana | | 5,0000 |
| Moda | | 5,00 |

Para el cálculo de la Media Geométrica, Media Armónica, con los mismos datos ya ingresados en el menú del SPSS escoger: ANALIZAR-INFORMES-RESUMENES DE CASOS- pasar la variable Peso_recien_nacido1 a la casilla de la derecha -ESTADISTICOS y seleccionar media armónica, media geométrica y pasar a la casilla de la derecha-CONTINUAR-ACEPTAR



**Resúmenes de casos**

Peso recien nacido1

| Media armónica | Media geométrica |
|---|---|
| 3,2949 | 3,8083 |

3.8. Ejercicios Propuestos

1. Los siguientes datos corresponden a pesos ( Kg) de 20 alumnos.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 50.5 | 54.5 | 52.4 | 54.5 | 55.0 | 61.5 | 61.0 | 60.0 | 62.0 | 58.5 |
| 51.4 | 52.5 | 62.5 | 50.5 | 50.0 | 50.5 | 52.5 | 58.5 | 64.2 | 50.5 |

a. Hallar e interprete la media aritmética, mediana y moda.

b. Si los pesos de los alumnos se incrementan en un 15 % mas ¾ Kg. Hallar el nuevo peso promedio y el peso mediano.

2. La siguiente tabla de distribución de frecuencias proporciona los montos por venta en cientos de nuevos soles de un grifo.

| Montos por Ventas | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 |
|---|---|---|---|---|---|---|
| $f_i$ | 12 | 15 | 20 | 25 | 12 | 8 |

a. Hallar e interprete la media aritmética, mediana y moda.
b. Si los montos por ventas se incrementan en 12%. Hallar el monto promedio y el monto mediano.

3. La tabla de distribución de frecuencias tiene como variable en estudio las estaturas (mts) de 50 personas.

| Estatura (mts) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 1.67-1.69 | 5 | | | | |
| 1.69-1.71 | 9 | | | | |
| 1.71-1.73 | 12 | | | | |
| 1.73-1.75 | 10 | | | | |
| 1.75-1.77 | 9 | | | | |
| 1.77-1.79 | 5 | | | | |
| Total | 50 | | | | |

Calcular e interprete:

a. $f_3$ ; $h_5$; $F_2$; $H_3$; $H_5 - H_2$; $x_3'$

b. La estatura media, mediana y modal de las personas.

4. El promedio de 10 calificaciones del curso de estadística en un salón de clase es de 9.8; pero al ingresar los datos a la computadora en lugar de digitar 14 se digita 4, y en lugar de digitar 8 digita 12. Calcular la nueva nota promedio con estas correcciones.

5. De la siguiente tabla de distribución de frecuencias Hallar la media aritmética, mediana, media geométrica y media armónica.

| Montos por ventas en | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X_i'$ |
|---|---|---|---|---|---|
| 18.5-23.5 | 5 | | | | |
| 23.5-28.5 | 11 | | | | |
| 28.5-33.5 | 20 | | | | |
| 33.5-38.5 | 10 | | | | |
| 38.5-43.5 | 4 | | | | |
| | 50 | | | | |

6. Un centro de información realizó durante los últimos 4 años las siguientes adquisiciones de revistas científicas: 30, 25, 20, 15, si el presupuesto para la adquisición se ha mantenido durante los 4 años. Hallar el promedio de revistas adquiridas por año.

7. La media aritmética de dos números es 10 y su media geométrica 8, Hallar su media armónica.

8. La población de un país en el año 1990 fue de 23,5 millones de habitantes, en el año 1995 fue de 25.6 millones de habitantes y en el año 2000 de 27.9 millones de habitantes. Hallar la población media para esos tres años.
9. Un centro de investigación ha experimentado un incrementado en la producción de revistas científicas tal como se detalla: 12% en el año 2000, 8% en el año 2001 y 14 % en 2002. Hallar el promedio de crecimiento anual.

10. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio al número de hijos.

| número de hijos | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|
| 1 | | 0.1 | | |
| 2 | | 0.2 | | |
| 3 | 28 | | | |
| 4 | 14 | | | |
| 5 | 7 | | | |
| Total | | | | |

Calcular:

$$\overline{x}, \quad me, \quad m_o$$

11. Los siguientes datos corresponden a edades (años) de alumnos de instrucción primaria.
12 14 12 12 13 14 15 12 14 14

Hallar la media, mediana y moda de las edades de los alumnos.

12. De la pregunta anterior, calcular la media, mediana y moda después de 5 años.
13. Los datos de la tabla, tiene como variable en estudio la edad de estudiantes(años)

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i^{'}$ |
|---|---|---|---|---|---|
| | | | | | 14 |
| | | 0.4 | | 0.56 | 18 |
| | 12 | | | | |
| | 8 | | | | |
| | 2 | | | | |
| Total | | | | | |

a. Completar la tabla de distribución de frecuencias
b. Hallar la media, mediana y moda de las edades de los estudiantes.

14. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio al número de hijos.

| número de hijos | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|
| 1 | | 0.1 | | |
| 2 | | 0.2 | | |
| 3 | | 0.4 | | |
| 4 | | 0.2 | | |
| 5 | | 0.1 | | |
| Total | | | | |

Hallar la media, mediana y moda de la variable número de hijos

15. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio a los montos por ventas registrados por una tienda comercial.

| Montos por ventas ( S/.) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 4,500 - 6,000 | | 0.15 | | | |
| 6,000 - 7,500 | | 0.20 | | | |
| 7,500 - 9,000 | | 0.45 | | | |
| 9,000 - 10,500 | | 0.12 | | | |
| 10,500 - 12,000 | | 0.08 | | | |
| Total | | | | | |

Hallar la media, mediana y moda de los montos por ventas.

16. La tabla contiene datos de los tiempos(minutos) registrados en atender a los usuarios que concurren a un centro de información:

| X=Tiempo (minutos) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 0.5 - 1.0 | | 0.10 | | | |
| 1.0 - 1.5 | | "a" | | | |
| 1.5 - 2.0 | | 0.35 | | | |
| 2.0 - 2.5 | | "b" | | | |
| 2.5 - 3.0 | | 0.1 | | | |
| Total | | | | | |

a. Si la media aritmética es 1.75, completar la tabla.
b. Hallar la moda y mediana del tiempo de atención a los usuarios que concurren al centro de información.

17. El valor de la mediana para un conjunto par de datos es 20. Si los datos se incrementan en 20% más 4 unidades. Hallar el nuevo valor de la mediana.

18. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio la edad (años) del jefe de familia.

| Edades (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 24 - 28 | 5 | | | | |
| 28 - 32 | 20 | | | | |
| 32 - 36 | 10 | | | | |
| 36 - 40 | 20 | | | | |
| 40 - 44 | 5 | | | | |
| Total | 60 | | | | |

Hallar: la media, mediana, moda de las edades del jefe de familia.

19. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio al número de artículos por revistas científicas.

| Número de artículos | $f_i$ | $h_i$ | $F_i$ | $H_i$ |
|---|---|---|---|---|
| 15 | 12 | | | |
| 9 | 8 | | | |
| 10 | 15 | | | |
| 14 | 18 | | | |
| 18 | 7 | | | |
| Total | 60 | | | |

Hallar la media, mediana moda del número de artículos por revistas.

# CAPÍTULO 4
# MEDIDAS DE VARIABILIDAD O DISPERSIÓN

**INTRODUCCIÓN**

En este CAPÍTULO se estudiarán las medidas estadística de variabilidad o dispersión para un conjunto de datos; si los datos fuesen iguales, se dice que no existe variabilidad o dispersión entre ellos, en caso contrario se dice que los datos presentan dispersión o variabilidad. El estudio de la variabilidad de un conjunto de datos es de suma importancia en todo análisis estadístico ya que de esto depende el grado de confiabilidad de las estadísticas que se calculan..

Las medidas de variabilidad o dispersión se clasifican en:

- MEDIDAS DE VARIABILIDAD ABSOLUTAS

Cuando su valor esta expresado en la misma unidad de medida de los datos originales. Las principales medidas de variabilidad absolutas son:

1. Rango
2. Variancia o Varianza
3. Desviación estándar

- MEDIDAS DE VARIABILIDAD RELATIVA

Cuando su valor carece de unidad de medida; la medida de variabilidad relativa que se usará es: El Coeficiente de variación o variabilidad

## 4.1. EL RANGO

El rango es una medida (distancia), a través de la cual se distribuyen todos los datos de la muestra o población. Se calcula por la diferencia entre el dato mayor y dato menor que toma la variable. Un valor del rango diferente de cero nos indica la presencia de variabilidad de los datos, mas no nos indica acerca de la variabilidad interna entre los datos.

Para datos sin agrupar: $Rango = Dato\ mayor - Dato\ \ menor$

Para datos agrupados: $Rango = LS_k - LI_1$

Donde:

$LS_k$ = límite superior de la última clase

$LI_1$ = límite inferior de la primera clase

Ejemplo:
Sean los montos por ventas de una tienda comercial expresada en nuevos soles:

| | | | | | |
|---|---|---|---|---|---|
| 1500 | 2300 | 1800 | 2000 | 1500 | 2400 |
| 2300 | 1000 | 1200 | 2400 | 2500 | 1800 |

De acuerdo con la definición el rango es:

$Rango = R = 2500 - 1000 = 1500$ Soles

Nos indica que 1500 nuevos soles es la distancia entre la venta menor y la mayor; el rango es una medida de dispersión de primera vista para la variabilidad de los datos; pero no evalúa el grado de variabilidad de los datos intermedios.

NOTA:

Si todos los datos fuesen iguales el rango es cero, nos indica que todos los datos se concentran en un mismo punto. Si el rango es diferente de cero nos indica que los datos no se concentran en un mismo punto; es decir existe variabilidad o dispersión de los datos

El rango puede ser positivo, negativo o cero
El rango posee unidad de medida, que es la misma de los datos originales.
El rango es bien sensible a la presencia de datos extremos (datos muy altos o datos muy bajo de lo normal)

## 4.2. LA VARIANCIA O VARIANZA

Su valor corresponde al promedio de las desviaciones al cuadrado de los datos con respecto a la media aritmética. Es decir su valor da una idea del grado de variabilidad de los datos con respecto a la media aritmética. La variancia para un conjunto de datos esta expresada en unidades cuadráticas. Por ejemplo, si la unidad de medida de los datos es kilogramos, la variancia será expresada en kilogramos al cuadrado. Este hecho dificulta la interpretación de la variancia.

- CÁLCULO DE LA VARIANCIA PARA LA POBLACIÓN PARA DATOS SIN AGRUPAR Y DATOS AGRUPADOS

Sea $X$ una variable en estudio que toma los valores $x_1, x_2, ..., x_N$, y $\mu$ su media aritmética poblacional, la variancia poblacional se simboliza por $\sigma^2$ y es definida por:

PARA DATOS SIN AGRUPAR

$$\sigma_x^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{\sum_{i=1}^{N} x_i^2 - N\mu^2}{N}$$

PARA DATOS AGRUPADOS

$$\sigma_x^2 = \frac{\sum_{i=1}^{k}(x'_i - \mu)^2 f_i}{N} = \frac{\sum_{i=1}^{N}(x'_i)^2 f_i - N\mu^2}{N}$$

- CÁLCULO DE LA VARIANCIA PARA UNA MUESTRA PARA DATOS SIN AGRUPAR Y AGRUPADOS

PARA DATOS SIN AGRUPAR

Sea $X$ una variable en estudio que toma los valores $x_1, x_2, ..., x_n$, y $\bar{x}$ su media aritmética muestral, la variancia para la muestra se simboliza por $S_x^2$ y es definida por:

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{X}^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}$$

PARA DATOS AGRUPADOS:

$$S_x^2 = \frac{\sum_{i=1}^{k}(x'_i - \bar{X})^2 f_i}{n-1} = \frac{\sum_{i=1}^{k}(x'_i)^2 f_i - n\bar{X}^2}{n-1} = \frac{\sum_{i=1}^{k}(x'_i)^2 f_i - \frac{(\sum_{i=1}^{k} x'_i f_i)^2}{n}}{n-1}$$

NOTA-1

Si los datos fuesen todos iguales la variancia es cero, esto nos indica que todos los datos están concentrados en la media aritmética, en otras palabras no existe variabilidad de los datos con respecto a la media aritmética.
Si los datos son diferentes, de acuerdo a la definición el valor de la variancia será mayor que cero, en otras palabras existe variabilidad entre los datos con respecto a la media aritmética.

NOTA-2

Es riesgoso usar el valor de la variancia para concluir que los datos de la muestra son muy o poco variable. Su uso es más que todo para comparar la variabilidad entre dos o más conjuntos de datos que tienen media aritmética similares.

Ejemplo-1

Sea una muestra aleatoria de los pesos kg de 15 alumnos de un salón de clases, Hallar la variancia.

35.5 37.5 38.0 38.0 39.0 37.0 36.8 35.8 36.0 35.0 35.0

36.5 38.5 38.0 38.0

De los datos, calculamos:

$$\sum_{i=1}^{15} x_i = 554.6$$

$$\sum_{i=1}^{15} x_i^2 = 20528.88$$

$$S^2 = \frac{\sum_{i=1}^{15} x_i^2 - \frac{(\sum_{i=1}^{15} x_i)^2}{15}}{15-1} = \frac{20528-88- \frac{(554.6)^2}{15}}{14} = 1.67638$$

$$S^2 = 1.67638 Kg^2$$

1.67638 $Kg^2$, este valor nos indica la presencia de variabilidad de los pesos de los alumnos con respecto al peso promedio.

Ejemplo-2

| $i$ | $pesos$(kg) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $X'_i$ |
|---|---|---|---|---|---|---|
| 1 | $[1.4-1.8)$ | 2 | 5.88% | 2 | 5.88% | 1.60 |
| 2 | $[1.8-2.2)$ | 4 | 11.76% | 6 | 17.64% | 2.0 |
| 3 | $[2.2-2.6)$ | 4 | 11.76% | 10 | 29.40% | 2.4 |
| 4 | $[2.6-3.0)$ | 5 | 14.71% | 15 | 44.11% | 2.8 |
| 5 | $[3.0-3.4)$ | 10 | 29.42% | 25 | 73.53% | 3.2 |
| 6 | $[3.4-3.8)$ | 7 | 20.59% | 32 | 94.12% | 3.6 |
| 7 | $[3.8-4.2)$ | 2 | 5.88% | 34 | 100% | 4.0 |
| TOTAL | | 34 | 100.0% | | | |

Calcular el valor de la variancia

$n = 34$

$$\sum_{i=1}^{15} (x'_i)^2 f_i = 308.48$$

$$\sum_{i=1}^{15} x_i f_i = 100$$

$$S^2 = \frac{\sum_{i=1}^{7} (x'_i)^2 - \frac{(\sum_{i=1}^{7} x'_i f_i)^2}{34}}{34-1} = \frac{308.48 - \frac{(100)^2}{34}}{33} = 0.43522$$

0.43522 $Kg^2$ , es el resultado del promedio de las desviaciones al cuadrado; este valor nos indica la presencia de variabilidad de los pesos de los recién nacidos con respecto a su peso promedio en $Kg^2$.

- PROPIEDADES DE LA VARIANCIA

1. Si los datos de la muestra son iguales; es decir los valores que toma los datos es constante; entones su variancia o variabilidad es igual a nula (cero); es decir:

$$x_1 = x_2 = ... = x_i = ... = x_n = m \rightarrow \overline{X} = m \rightarrow S^2 = 0$$

2. Cuando a todos los datos de la muestra, se le multiplica por una constante "$a$" y además se le suma o resta otra constante "$b$", entonces la variancia de los datos transformados, es igual a la primera constante al cuadrado multiplicado por la variancia de los datos originales.

$x_1, x_2, ..., x_i ..., x_n \rightarrow \overline{X} \rightarrow S_x^2$, sea la transformación $y_i = ax_i \pm b$ , luego se cumple que la variancia de los datos transformados es igual a:

$$S_y^2 = a^2 S_x^2$$

4.3. DESVIACIÓN ESTANDAR

Su valor, se define como la raíz cuadrada positiva de la variancia; es decir,

$$S_x = \sqrt{Variancia}$$

El valor de la desviación estándar para un conjunto de datos esta expresado en las mismas unidades de la variable en estudio. Su valor es mayor o igual a cero, un valor cero indica que los datos de la muestra esta concentrados en un mismo punto ( no hay variabilidad o dispersión), un valor mayor a cero indica la presencia de variabilidad o dispersión de los datos con respecto a la media aritmética.

Ejemplo:
De los ejemplos anteriores se tiene:

Para datos sin agrupar:

$$S_x = \sqrt{1.67638kg^2} = 1.29475Kg$$

Como la media aritmética de los datos es 36.97 kg., y la desviación estándar 1.29 kg. Valor que indica la presencia variabilidad o dispersión de los datos con respecto a la media aritmética.

Para datos agrupados:

$$S_x = \sqrt{0.43522 kg^2} = 0.65971 Kg$$

Interpretación

1.29475 $Kg$ , es la desviación promedio con respecto a la media aritmética.
0.65971 $Kg$ , es la desviación promedio con respecto a la media aritmética.
Ambos valores obtenidos nos indica la presencia de variabilidad de los datos con respecto a la media.

NOTA-1

La desviación estándar al igual que la variancia, se utiliza para comparar la variabilidad o el grado de dispersión de dos o más conjunto de datos que poseen la misma unidad de medida y tienen sus promedios similares; es decir si se tiene dos a más conjunto de datos que poseen la misma unidad de medida y promedio similares, tendrá mayor dispersión o mayor variabilidad con respecto a la media aritmética aquel conjunto de datos que tiene la mayor variancia o desviación estándar.

Ejemplo

De la tabla que conjunto de datos son mas variables o están mas dispersos

| Muestras | Datos | Media Aritmética | Variancia | Desviación estándar |
|---|---|---|---|---|
| 1 | 4, 5, 6 | 5 | 1 | 1 |
| 2 | 1, 5, 9 | 5 | 16 | 4 |

Los datos de la muestra 2 están mas dispersos que la muestra 1

## 4.4. COEFICIENTE DE VARIACIÓN

Es una medida de variación relativa no tiene unidad de medida se calcula haciendo uso de la siguiente formula:

$$CV = \frac{S}{\overline{X}} x100\%$$

El valor de coeficiente de variación $CV$ , indica el número de veces que la desviación estándar contiene a la media aritmética, cuando contiene más veces a la media indica que los datos son más variables. Su valor se usa para comparar la variabilidad o dispersión entre dos o más conjuntos de datos que poseen diferentes unidades de medidas y medias aritméticas diferentes.

Sean A y B dos conjuntos de datos si el coeficiente de variación para el conjunto A es mayor al coeficiente de variación para B, entonces los datos de A presentan mayor variabilidad o dispersión relativa que los datos de B.

Ejemplo-1

Sean los siguientes datos:

1, 2, 3, 4, 5

Hallar e interprete el coeficiente de variación

$\overline{x} = 3$

$S^2 = 2.5$

$S = 1.5811$

$CV = \frac{1.5877}{3} x100\% = 52.70\%$ (Nos indica que la desviación estándar para este conjunto de datos representa el 52.70% de su media aritmética)

Ejemplo-2

La siguiente información corresponden a los montos (S/.) por ventas de dos tiendas comerciales A y B.

| Valores estadísticos | Tienda Comercial A | Tienda Comercial B |
|---|---|---|
| Venta promedio | 22,000 | 18,500 |
| Desviación estándar | 2640 | 3145 |

¿Qué montos por ventas están más dispersos?

Para la tienda Comercial A

$$CV = \frac{2640}{22000} x100\% = 12.0\%$$

Para la tienda Comercial B

$$CV = \frac{3145}{18500} x100\% = 17.0\%$$

Por lo tanto los montos por ventas de la tienda comercial B son más variables o más dispersos que los montos por ventas de A.

Aplicando el SPSS para datos no agrupados Media, Rango, Varianza, Desviación estándar,

Definiendo las variables

*Sin título1 [ConjuntoDatos0] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

| | Nombre | Tipo | Anchura | Decimales | Etiqueta | Valores | Perdidos | Columnas | Alineación | Medida | Rol |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Peso_recie... | Numérico | 8 | 1 | Peso recien na... | Ninguna | Ninguna | 8 | Derecha | Escala | Entrada |

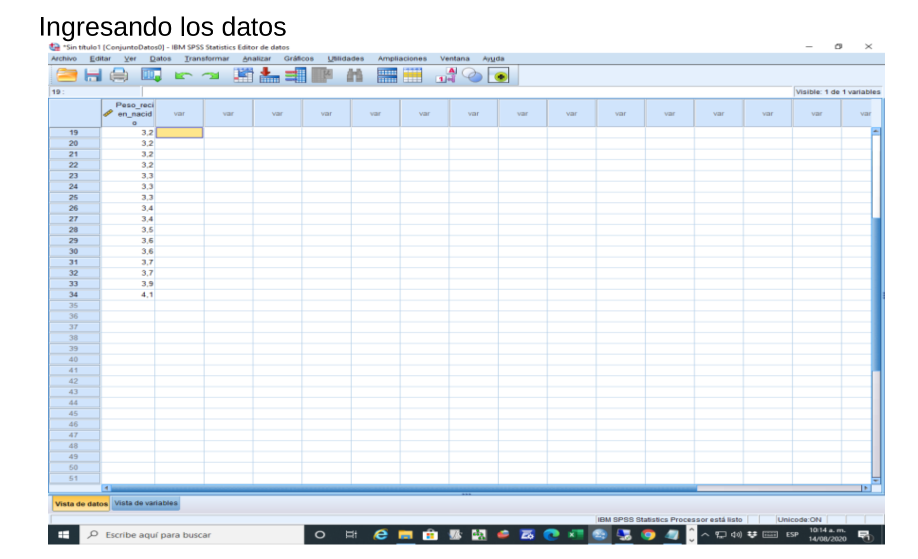
Vista de datos | Vista de variables

IBM SPSS Statistics Processor está listo | Unicode:ON

Escribe aquí para buscar | ESP | 10:12 a. m. 14/08/2020

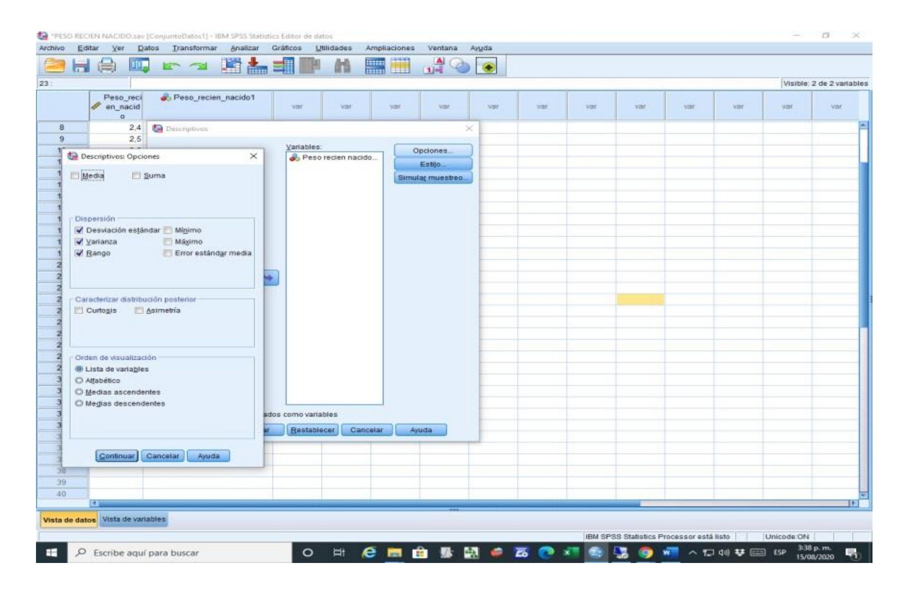
7

Ingresando los datos




Calculando el Rango, la Varianza, la Desviación estándar,

En el Menú del SPSS escoger ANALIZAR-ESTADISTICOS DESCRIPTIVOS--pasar la variable Peso_recien_nacido a la casilla de la derecha-OPCIONES- Desviación estandar; Varianza, Rango-CONTINUAR-ACEPTAR



**Estadísticos descriptivos**

| | N | Rango | Desv. Desviación | Varianza |
|---|---|---|---|---|
| Peso recien nacido | 34 | 2,6 | ,6600 | ,436 |
| N válido (por lista) | 34 | | | |

Aplicando el SPSS para datos agrupados: Rango, Varianza, Desviación estándar, con datos ya ingresados; en el menú del SPSS escoger:

Para datos agrupados
Se elabora la tabla de frecuencias para datos agrupados:

7. Cálculo del Rango” R “.
   R = peso mayor –peso menor.

   R = 4.1-1.5 = 2.6

8. Cálculo del número de clases (K).

$$K = 1 + 3.3Log(34) = 6.05388$$

   Como el valor que toma K debe ser un número entero, usar el redondeo simple (redondeo por defecto), se tiene que K = 6.

   OBSERVACIÓN
   El redondeo simple significa, que si un número tiene como primer decimal un número mayor o igual a 5, entonces la cifra entera del número se incrementa en una unidad.

   Ejemplo:
   Si, K = 6.7, usando el redondeo simple K = 7;
   Si, K = 6.4, usando el redondeo simple K = 6;

9. Cálculo del ancho del intervalo de clase (A)
   El ancho del intervalo de clase se obtiene a través de la siguiente fórmula:

$$A = \frac{R}{K}$$

   Para los datos se tiene:

$$A = \frac{R}{K} = \frac{2.6}{6} = 0.43333$$

El valor que toma “A”, debe tener el mismo número de decimales de los datos originales. Por redondeo simple se tiene:

$$A = 0.4$$

$$E = AK - R = 0.4(6) - 2.6 = -0.2$$

Como el error es negativo, se debe incrementar en una clase más(k=7)

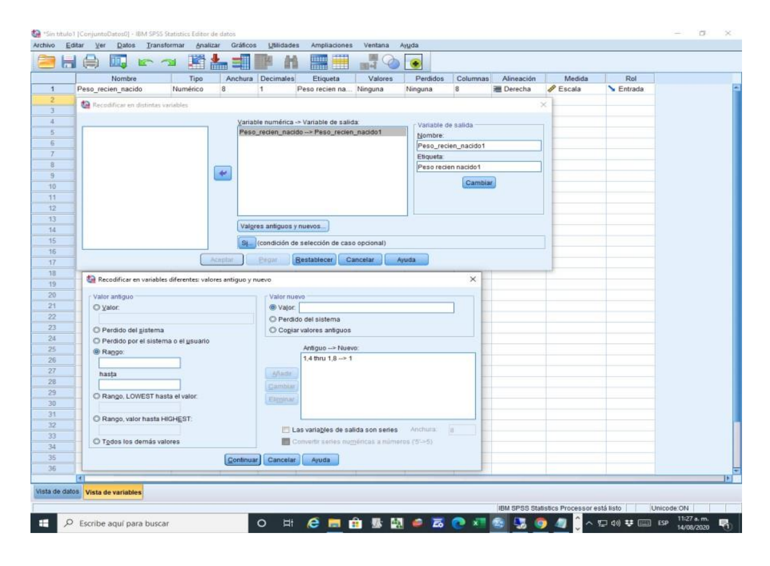
$$A = \frac{2.6}{7} = 0.4$$

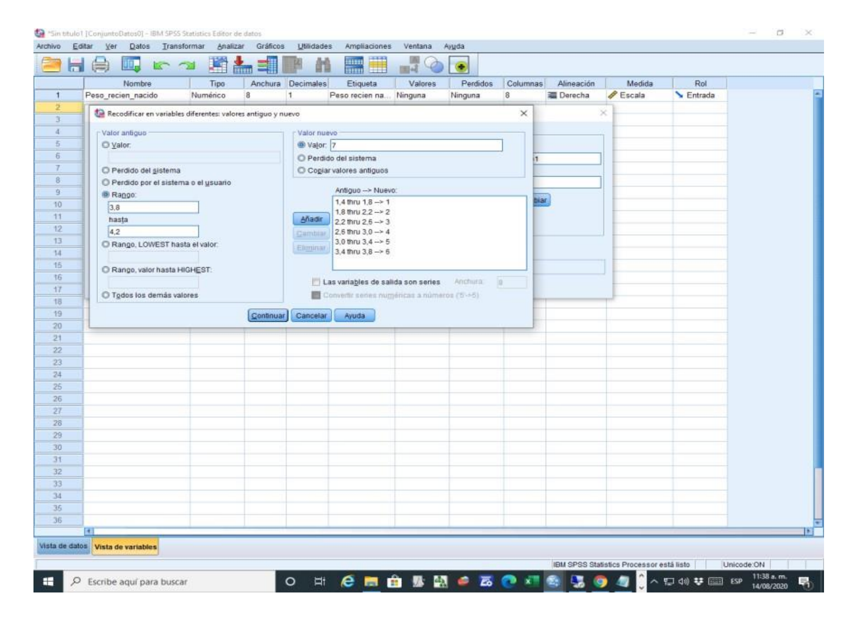
$$E = AK - R = 0.4(7) - 2.6 = 0.2$$

Cuando el error es positivo, la tabla tendrá 7 clases y ancho 0.4

$$valor\ de\ arranque = dato\ menor - \frac{E}{2} = 1.5 - \frac{0.2}{2} = 1.4$$

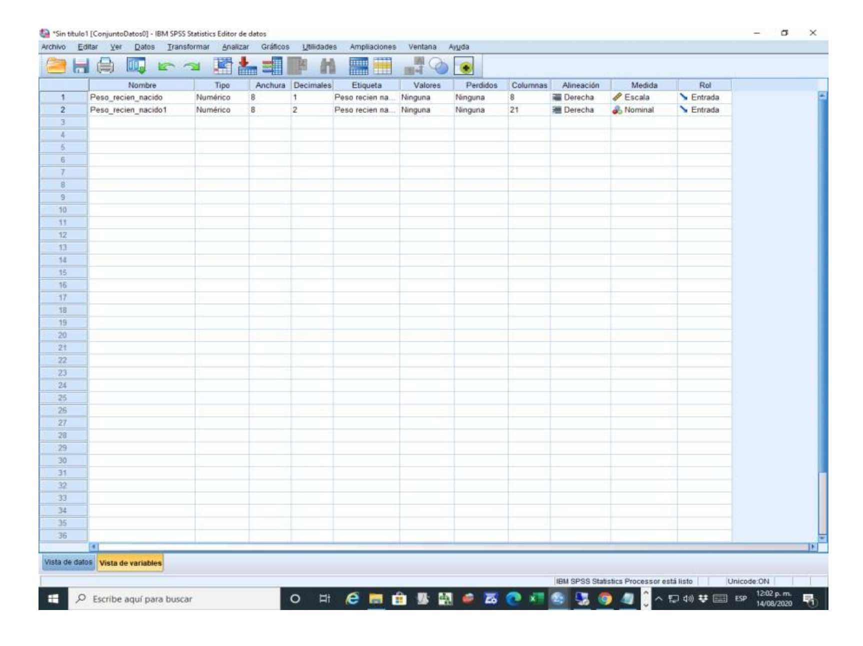
En el menú del SPSS, escoger la opción TRANSFORMAR-RECODIFICAR EN DISTINTAS VARIABLES-pasar la variable peso-recién-nacido al casillero de la derecha, en la derecha en donde indica variable de salida Nombre: poner peso_recien_nacido1 y en la etiqueta: poner Peso_recien_nacido1 – PRESIONAL LA OPCION CAMBIAR-VALORES ANTIGUOS Y NUEVOS-RANGO-poner desde 1,4 hasta 1,8, luego en VALOR poner 1 y AÑADIR, así sucesivamente continua con todos los rangos de 3,8 hasta 4,2.



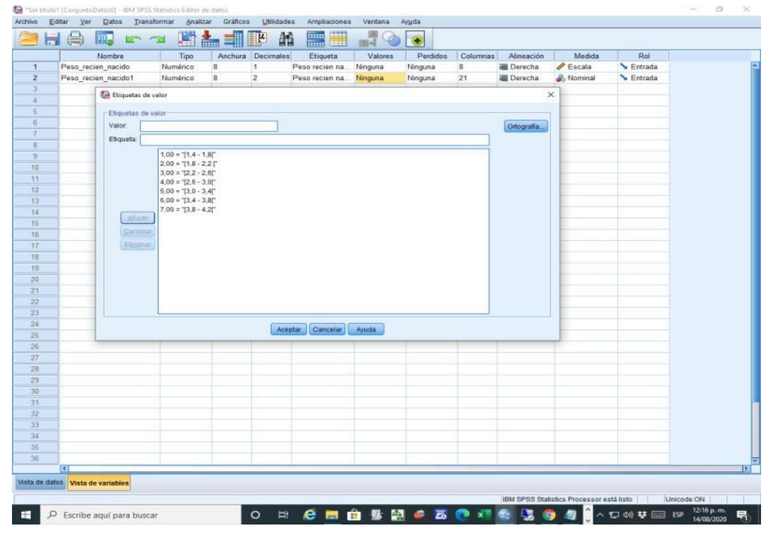



Cuando se termine de ingresar los rangos presionar, continuar y aceptar.

Y se crea una variable Peso_recien_nacido_1.

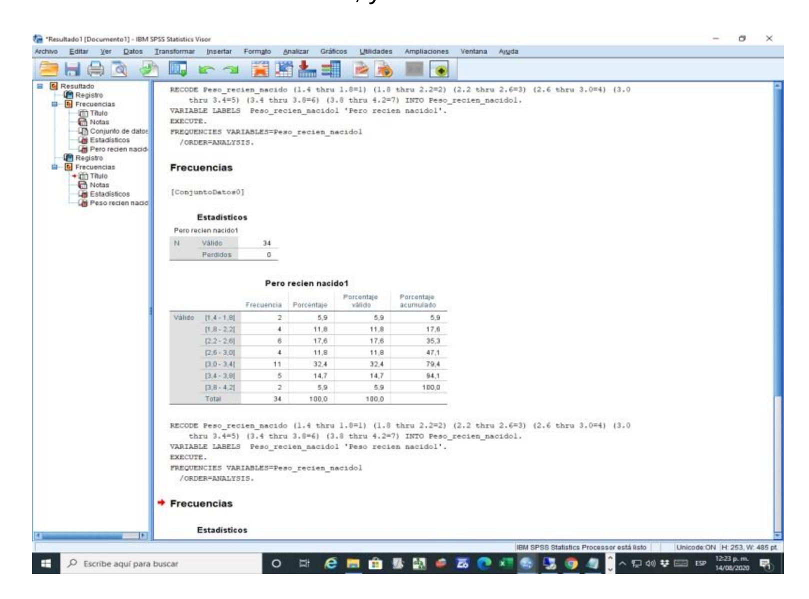


En la variable Peso_recien_nacido_1, presionar la opción valores e ingresar:



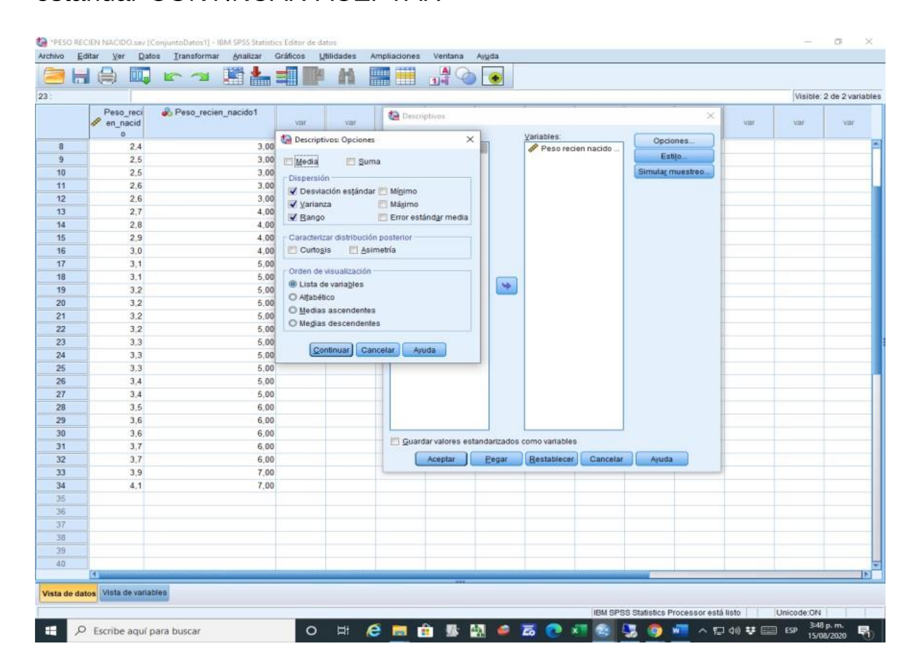

Y Poner aceptar.

En el Menú del SPSS escoger la opción ANALIZAR-ESTADISTICOS DESCRIPTIVOS-FRECUENCIA-para la variable Peso_recien_nacido 1 al casillero derecho-ACEPTAR, y nos muestra los resultados



Calculando el Rango, la Varianza, la Desviación estándar

En el Menú del SPSS escoger ANALIZAR-ESTADISTICOS DESCRIPTIVOS- -pasar la variable Peso_recien_nacido1 a la casilla de la derecha-OPCIONES-El Rango, La Varianza, la Desviación estándar-CONTINUAR-ACEPTAR




**Estadísticos descriptivos**

| | N | Rango | Desv. Desviación | Varianza |
|---|---|---|---|---|
| Peso recien nacido1 | 34 | 6,00 | 1,62889 | 2,653 |
| N válido (por lista) | 34 | | | |

4.5. <u>EJERCICIOS PROPUESTOS</u>

1. Los siguientes datos corresponden a estaturas (mts) de los alumnos de cierta universidad.

| | | | | | | |
|---|---|---|---|---|---|---|
| 1.71 | 1.68 | 1.85 | 1.58 | 1.68 | 1.66 | 1.72 |
| 1.67 | 1.68 | 1.65 | 1.68 | 1.65 | 1.68 | 1.67 |
| 1.70 | 1.70 | 1.68 | 1.65 | 1.66 | 1.68 | |

Hallar e interprete las medidas de dispersión absoluta y relativa

2. Los siguientes datos, corresponden a una muestra al azar de 40 frascos conteniendo mermelada de piña, los pesos (grs) se presentan a continuación:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 149 | 164 | 171 | 176 | 181 | 185 | 187 | 190 |
| 156 | 165 | 172 | 176 | 182 | 186 | 187 | 194 |
| 157 | 165 | 174 | 178 | 183 | 186 | 188 | 196 |
| 158 | 168 | 175 | 180 | 183 | 188 | 188 | 198 |
| 162 | 170 | 175 | 180 | 184 | 187 | 189 | 202 |

a. Elaborar la tabla de distribución de frecuencias.
b. Calcular las medidas de dispersión o variabilidad.

3. Se eligen al azar a 70 trabajadores de una empresa, la distribución de los ingresos mensuales se presentan en la tabla siguiente.

| $i$ | Ingresos | $f_i$ |
|---|---|---|
| 1 | $[400 \quad 600)$ | 6 |
| 2 | $[600 \quad 800)$ | 12 |
| 3 | $[800 \quad 1000)$ | 18 |
| 4 | $[1000 \quad 1200)$ | 15 |
| 5 | $[1200 \quad 1400)$ | 12 |
| 6 | $[1400 \quad 1600)$ | 7 |
| | TOTAL | 70 |

Usando la información de la tabla, calcular:
a. El Rango
b. La variancia
c. La desviación estándar.
d. El coeficiente de variación
e. la variancia y desviación estándar.

f. La variancia y desviación estándar, si la empresa decide aumentar los ingresos de sus trabajadores en un 7% más 45 nuevos soles.

***4.*** Una agencia bancaria, con la finalidad de reducir el tiempo en la atención a los clientes, elige una muestra aleatoria de 60 clientes en un día determinado y registra el tiempo (minutos) que demora en ser atendido, estos resultados se presentan en la tabla**.**

| $i$ | Tiempos | $h_i$ |
|---|---|---|
| 1 | $[0.5\ 1.0)$ | 0.1 |
| 2 | $[1.0\ 1.5)$ | $a$ |
| 3 | $[1.5\ 2.0)$ | 0.3 |
| 4 | $[2.0\ 2.5)$ | $b$ |
| 5 | $[2.5\ 3.0)$ | 0.15 |
| | TOTAL | 1 |

a. Calcular el valor de "a" y "b" , si se sabe que el tiempo promedio de atención a un cliente es de 1.825.

b. Calcular la desviación estándar.

c. Calcular el coeficiente de variación.

6. Los siguientes datos muestran los montos por ventas, en miles de dólares, de 20 vendedores de una compañía de computadora.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 40.2 | 29.3 | 35.6 | 88.2 | 42.9 | 26.9 | 28.7 | 99.8 | 35.6 | 37.8 |
| 44.2 | 32.3 | 55.2 | 50.6 | 25.4 | 31.7 | 36.8 | 45.2 | 25.1 | 39.7 |

a. Calcular la variancia, desviación estándar y coeficiente de variación.

b. Si las ventas se incrementan en un 15% con respecto a los montos iniciales. Calcular el rango, la variancia, desviación estándar y coeficiente de variación.

c. Elaborar la tabla de distribución de frecuencias. Calcular: rango, variancia, desviación estándar y coeficiente de variación.

7. Una muestra aleatoria de tamaño 20 tiene por media 5 y variancia igual a 0.36, sean las siguientes transformaciones:

   a. Si a cada dato se le incrementa en un 20% de su valor más 0.5 unidades.
   b. Si a cada dato se le incrementa en 25% de su valor más 0.2 unidades.

¿Que conjunto de los nuevos datos presenta la mayor variación?

Ejercicios adicionales

1. Sea la Tabla de datos:

| $i$ | $X$ | $Y$ | $Z$ |
|---|---|---|---|
| 1 | 2 | 1 | 1 |
| 2 | 1 | 1 | 1 |
| 3 | 1 | 2 | 3 |
| 4 | 3 | 2 | 1 |
| 5 | 1 | 2 | 1 |

Hallar el valor de:

a. $$\sum_{i=1}^{5}(x_i + y_i)(x_i + z_i)(y_i + z_i)$$

b. $$\sum_{i=1}^{5}(x_1^2 + y_2^2 + z_3^2 + x_i y_i + x_i z_i + y_i z_i)^2$$

2. Sea la tabla de datos

| $i$ | $x_i$ | $f_i$ |
|---|---|---|
| 1 | 10 | 2 |
| 2 | 11 | 4 |
| 3 | 12 | 6 |
| 4 | 13 | 4 |
| 5 | 14 | 2 |
| | TOTAL | 18 |

Hallar el valor de:

$$\sum_{i=1}^{5}(x_i - \bar{x})^2 f_i \quad siendo \quad \bar{x} = \frac{\sum_{i=1}^{5} x_i \,.\, f_i}{5}$$

3. Una muestra tomada al azar de 100 estudiantes que concurren a la Biblioteca Central de la Universidad Nacional Mayor de San Marcos, arroja que: el 70% son varones; la edad promedio es de 28.4 años; el 15% pertenecen a la Escuela Académico Profesional de Bibliotecología; el 70% opinan que los servicios que presta la biblioteca es bueno; el gasto promedio mensual es de S/.450.00. Con la información, establecer la población en estudio, la unidad elemental, tamaño de la muestra, variables, tipos de variables, parámetros y valores estadísticos.

4. Una muestra aleatoria de 84 datos, arroja que el dato menor es 17.1 y el dato mayor 37.2. Determinar los intervalos de clases para la tabla de distribución de frecuencias.

5. Sea la tabla de distribución de frecuencias

| Edades(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| [ 20 > | | | 4 | | 18 |
| [ > | | | 12 | | |
| [ > | | | 28 | | |
| [ > | | | 36 | | |
| [ > | | | 40 | | |
| Total | | | | | |

Completar la tabla

6. Sea la tabla de distribución de frecuencias, cuya variable en estudios es la edad (años) de usuarios, que concurren a un centro de información.

| Edades(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| [20 24 > | 4 | | | | |
| [24 28 > | 6 | | | | |
| [28 32 > | 15 | | | | |
| [32 36 > | 15 | | | | |
| [36 40 > | 6 | | | | |
| [40 44 > | 4 | | | | |
| Total | 50 | | | | |

a. Calcular e interpretar $h_i$, $H_5 - H_2$

b. ¿Qué cantidad de usuarios tienen edades superior a los 26 años?.

c. ¿Que % de usuarios tienen edades entre los 26 y 38 años.

7. Una encuesta realizada a 20 jefes de hogar arroja los siguientes datos.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 4 | 2 | 0 | 1 | 2 |
| 3 | 3 | 2 | 0 | 1 | 1 | 2 | 4 |
| 2 | 2 | 1 | 2 | | | | |

Elaborar la tabla de distribución de frecuencias y GRÁFICO respectivo.

8. Una muestra aleatoria de 90 datos, arroja que el dato menor es 20.45 y el dato mayor 70.35. Determinar los intervalos de clases para la tabla de distribución de frecuencias.

9. Sea la tabla de distribución de frecuencias

| Edades(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| [7 > | | 0.1 | | | 10 |
| [ > | 8 | | | | |
| [ > | | 0.4 | | | |
| [ > | | 0.2 | | | |
| [ > | 4 | | | | |
| Total | | | | | |

Completar la tabla

10. Sea la tabla de distribución de frecuencias, cuya variable en estudios es la edad (años) de usuarios, que concurren a un centro de información.

| Edades(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| [20 24 > | 4 | | | | |
| [24 28 > | 6 | | | | |
| [28 32 > | 15 | | | | |
| [32 36 > | 15 | | | | |
| [36 40 > | 6 | | | | |
| [40 44 > | 4 | | | | |
| Total | 50 | | | | |

a. Calcular e interpretar $h_5$, $H_6 - H_2$

b. ¿Qué cantidad de usuarios tienen edades inferiores a los 30 años?.

c. ¿Que % de usuarios tienen edades entre los 22 y 42 años.

11. Una encuesta realizada a 20 jefes de hogar arroja los siguientes datos.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 4 | 2 | 0 | 1 | 2 |
| 3 | 3 | 2 | 0 | 3 | 1 | 2 | 4 |
| 2 | 2 | 1 | 2 | | | | |

Elaborar la tabla de distribución de frecuencias y GRÁFICO respectivo.

12. La media aritmética de 20 datos es 4.5, si a cada dato se incrementa en su 20% más 3 unidades. Calcular su nueva media aritmética.

13. El peso menor y mayor de 80 datos son 58.2 kg y 83.2 kg respectivamente: Determinar las clases para la tabla de distribución de frecuencias.

14. Los siguientes datos corresponden a una encuesta tomada a 30 jefes de familias, en la cual la variable en estudios es el número de hijos por familia.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 1 | 2 | 3 | 2 | 1 | 2 | 2 | 4 |
| 5 | 3 | 3 | 4 | 3 | 3 | 2 | 2 | 3 | 4 |
| 4 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 3 | 5 |

Calcular la media aritmética del número de hijos por familia.

15. Distribución de las edades (años) de estudiantes de la UNMSM:

| Edades | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18-22 | 10 | | | | |
| 22-26 | | 0.1875 | | | |
| 26-30 | 30 | | | | |
| 30-34 | 15 | | | | |
| 34-38 | | 0.1250 | | | |
| TOTAL | | | | | |

a. Completar la tabla de distribución de frecuencias
b. Que % de estudiantes tienen edades entre los 20 y 32 años.
c. Calcular la media aritmética de las edades.

16. La media aritmética de 20 datos es 4.5, si a cada dato se incrementa en su 20% más 3 unidades. Calcular su nueva media aritmética.

17. Los siguientes datos corresponden a una encuesta tomada a 30 jefes de familias, en la cual la variable en estudios es el número de hijos por familia.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 1 | 2 | 3 | 2 | 1 | 2 | 2 | 4 |
| 5 | 3 | 3 | 4 | 3 | 3 | 2 | 2 | 3 | 4 |
| 4 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 3 | 5 |

Calcular la media aritmética, la mediana y la moda del número de hijos.

18. Distribución de las edades (años) de estudiantes de la UNMSM:

| Edades | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18 - 22 | | 0.1250 | | | |
| 22 - 26 | | 0.1875 | | | |
| 26 - 30 | | 0.3750 | | | |
| 30 - 34 | | 0.1875 | | | |
| 34 - 38 | | 0.1250 | | | |
| TOTAL | | 1.0000 | | | |

Calcular la media aritmética, la mediana y moda de las edades de los estudiantes.

19. Un estudiante del curso de estadística, calcula la media aritmética de 20 datos obteniendo como resultado 50; otro estudiante con los mismos datos obtiene por media aritmética

48, si el primer estudiante se equivoco al ingresar a la calculadora solo el último dato. Además la suma del dato errado y su verdadero es 100. Calcular el valor del dato verdadero.

20. La mediana de un conjunto de datos par es 35, si cada dato se incrementa en 15% más tres unidades, hallar el valor de la nueva mediana.

21. Los siguientes datos corresponden al número de artículos de una muestra elegidas al azar de 30 revistas científicas.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 12 | 14 | 11 | 12 | 13 | 12 | 11 | 12 | 12 | 14 |
| 15 | 13 | 13 | 14 | 13 | 13 | 12 | 12 | 13 | 14 |
| 14 | 13 | 13 | 14 | 14 | 13 | 13 | 13 | 13 | 15 |

Calcular la media aritmética, mediana y moda del número de artículos por revistas.

22. Distribución de las edades (años) de estudiantes de la UNMSM:

| Edades(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18 - 22 | | 0.10 | | | |
| 22 - 26 | | 0.45 | | | |
| 26 - 30 | | 0.30 | | | |
| 30 - 34 | | 0.11 | | | |
| 34 - 38 | | 0.04 | | | |
| TOTAL | | 1.00 | | | |

a. Calcular la media aritmética, mediana y moda de las edades de los estudiantes.
b. Calcular el 4to. decil y 3er. cuartil.
c. Que edades encierran al 70% central de los estudiantes.

23. Distribución de las edades de estudiantes:

| Edad(años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 16-20 | | 0.05 | | | |
| 20-24 | | 0.25 | | | |
| 24-28 | | 0.40 | | | |
| 28-32 | | 0.25 | | | |
| 32-36 | | 0.05 | | | |
| TOTAL | | 1.00 | | | |

Calcular para los montos por ventas:

a. Media aritmética
b. Mediana
c. Moda
d. El tercer decil
e. El primer cuartil
f. El 82 avo percentil
g. Calcular la curtosis, interpretación

24. Los siguientes datos corresponden a una encuesta tomada a 30 jefes de familias, en la cual la variable en estudios es el número de periódicos comprados durante una semana determinada.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 1 | 2 | 3 | 2 | 1 | 2 | 2 | 4 |
| 5 | 3 | 3 | 4 | 3 | 3 | 2 | 2 | 3 | 4 |
| 4 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 3 | 5 |

Elaborar la tabla de distribución de frecuencias y su GRÁFICO correspondiente para la variable en estudio.

25. Distribución de las edades (años) de estudiantes de la UNMSM:

| Edades | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | 10 | | | | 22 |
| | | 0.1875 | 25 | | 26 |
| | | | | 0.6875 | |
| | 15 | | | | |
| | | | | | |
| TOTAL | 80 | | | | |

a. Completar la tabla de distribución de frecuencias

b. Calcular e interpretar: $f_3$ , $F_3$, $H_4 - H_2$

c. Que % de estudiantes tienen edades inferiores a los 30 años.

26. La tabla contiene datos sobre la edad de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18-22 | | 0.10 | | | |
| 22-26 | | "a" | | | |
| 26-30 | | 0.35 | | | |
| 30-34 | | "b" | | | |
| 34-38 | | 0.1 | | | |
| Total | | | | | |

Si la media aritmética es 28.2 años.
Calcular, la mediana, la moda, el 4to decil, el tercer cuartil y el 82 percentil.

27. La variancia de un conjunto de datos no agrupados es 20, si cada dato se incrementa en un 10%. Hallar el nuevo valor de la variancia.

28. La estatura menor y mayor de 60 personas son 1.62 metros y 1.82 metros, determinar para la tabla de distribución de frecuencias.

29. La tabla contiene datos sobre la edad de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i^{'}$ |
|---|---|---|---|---|---|
| 20-24 | | 0.10 | | | |
| 24-28 | | "a" | | | |
| 28-32 | | 0.30 | | | |
| 32-36 | | "b" | | | |
| 36-40 | | 0.1 | | | |
| Total | | | | | |

Si la media aritmética es 30 años.
Calcular, la mediana, la moda, el 6to decil, el primer cuartil y el 82 percentil de las edades.

30. La media aritmética y variancia de 4 datos es 6 y $\frac{10}{3}$, respectivamente, una revisión de los datos, se encontró que en lugar de ingresar a la calculadora el valor 8 se ingreso por error el valor 4. Con esta información, calcular la media aritmética y variancia corregidas.

31. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio el número de artículos de 80 revistas científicas sobre biología elegidas al azar.

| X=N° de artículos por revistas | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i^{'}$ |
|---|---|---|---|---|---|
| 21 | 5 | | | | |
| 17 | 5 | | | | |
| 19 | 20 | | | | |
| 18 | 40 | | | | |
| 20 | 10 | | | | |
| Total | 80 | | | | |

Hallar: la media, mediana, moda, la variancia y el coeficiente de variación.

32. De una muestra de 80 datos el mayor y menor son 56.25 y 12.34. respectivamente. Determinar los intervalos de clases para la tabla de distribución de frecuencias.

33. La media de 20 datos es 5; la media de 30 datos es 6: Hallar la media de los 50 datos.

34. Sean 50 datos, tal que $\bar{x} = 2$, además $\sum_{i=1}^{50} (x_i + \bar{x})^2 = 1000$. Hallar su coeficiente de variación.

35. La tabla contiene datos sobre la edad de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | 0.10 | | | 22 |
| | 12 | | | | 26 |
| | | 0.20 | | | |
| | | 0.50 | | | |
| | 4 | | | | |
| Total | | | | | |

a. Completar la tabla

b. Hallar la media aritmética, mediana y moda

c. Que porcentajes de usuarios tienen edades entre 22 y los 30 años.

d. Hallar la variancia y el coeficiente de variabilidad.

36. La media aritmética de 20 datos es 5; la media aritmética de 30 datos es 6: Hallar la media aritmética de los 50 datos.

37. Sean 50 datos, tal que $\bar{x} = 4$, además $\sum_{i=1}^{50} x_i^2 = 849$. Hallar su coeficiente de variación.

38. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | 0.10 | | | 22 |
| | 12 | | | | 26 |
| | | 0.20 | | | |
| | | 0.50 | | | |
| | 4 | | | | |
| Total | | | | | |

a. Completar la tabla.

b. Hallar la media aritmética, mediana y moda.

c. Hallar la variancia y el coeficiente de variabilidad.
d. Hallar el coeficiente de asimétrica. Interpretar resultado

39. La media aritmética de 20 datos es 5; la media aritmética de 30 datos es 6. Hallar la media aritmética de los 50 datos.

40. La mediana de 60 datos es 18. Si a cada dato se incrementa en un 12% más 3 unidades. Hallar el nuevo valor de la mediana.

41. Una muestra tomada al azar de 50 revistas científicas en biología, proporciona los siguientes datos con respecto al número de artículos por revistas.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 12 | 12 | 12 | 13 | 13 | 13 | 13 |
| 13 | 13 | 13 | 13 | 14 | 14 | 14 | 14 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 14 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| 15 | 16 | 16 | 16 | 16 | 16 | 17 | 17 |
| 17 | 17 | 18 | 18 | 18 | 18 | 19 | 19 |
| 19 | 19 | | | | | | |

Calcular:

a. La media aritmética, la mediana, la moda del número de artículos por revistas.
b. El tercer decil, el tercer cuartil y el ochenta y cinco avo percentil del número de artículos por revistas.

42. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 20-24 | | 0.10 | | | |
| 24-28 | | 0.20 | | | |
| 28-32 | | 0.46 | | | |
| 32-36 | | 0.16 | | | |
| 36-40 | | 0.08 | | | |
| Total | | 1.00 | | | |

a. Hallar la media aritmética, mediana y moda de las edades de los usuarios.
b. Hallar la media aritmética de las edades de los usuarios 3 años después.

43. El menor de los pesos de 70 estudiantes de la UNMSM es 56.4 kg y el peso mayor es de 94.3 kg. Determinar las clases, para la tabla de distribución de frecuencias.

44. Una muestra tomada al azar de 50 revistas científicas en biología, proporciona los siguientes datos con respecto al número de artículos por revistas.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 12 | 12 | 12 | 12 | 13 | 13 | 13 |
| 13 | 13 | 13 | 13 | 13 | 13 | 13 | 14 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 14 | 14 | 14 | 15 | 15 | 15 | 15 | 15 |
| 15 | 15 | 15 | 15 | 15 | 16 | 16 | 16 |
| 16 | 16 | | | | | | |

Calcular:
La media aritmética, la mediana, la moda, el cuarto decil, el primer cuartil y el setenta y dos avo percentil del número de artículos por revistas.

45. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | 0.10 | | | 22 |
| | 20 | | | | |
| | | 0.46 | | | 30 |
| | 16 | | | | |
| | | 0.08 | | | |
| Total | | 1.00 | | | |

a. Completar la tabla de distribución de frecuencias.
b. Hallar la media aritmética, mediana y moda de las edades de los usuarios.
c. ¿Qué porcentaje de usuarios tienen edades entre los 26 y 34 años.
d. De la tabla calcular el valor de $\dfrac{\sum_{i=1}^{5}(x_i' - \bar{x})^2 . f_i}{n-1}$

46. El menor de los pesos de 70 estudiantes de la UNMSM es 57.4 kg y el peso mayor es de 93.3 kg. Determinar las clases, para la tabla de distribución de frecuencias.

47. Una muestra tomada al azar de 50 revistas científicas en biología, proporciona los siguientes datos con respecto al número de artículos por revistas.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 12 | 12 | 12 | 12 | 13 | 13 | 13 |
| 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| 13 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| 16 | 16 | | | | | | |

Calcular:
La media aritmética, la mediana, la moda, el cuarto decil, el primer cuartil y el setenta y dos avo percentil del número de artículos por revistas.

48. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| [ 22 > | | 0.10 | | | 20 |
| [ > | 20 | | | | |
| [ > | | 0.30 | | | |
| [ > | 8 | | | | |
| [ > | | 0.04 | | | |
| Total | | 1.00 | | | |

a. Completar la tabla de distribución de frecuencias.
b. Hallar la media aritmética, mediana y moda de las edades de los usuarios.
c. ¿Qué porcentaje de usuarios tienen edades entre los 24 y 32 años.
d. De la tabla calcular el valor de $\dfrac{\sum_{i=1}^{5}(x_i' - \bar{x})^2 . f_i}{n-1}$

49. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18 - 22 | | 0.06 | | | |
| 22 - 26 | | 0.50 | | | |
| 26 - 30 | | 0.20 | | | |
| 30 - 34 | | 0.12 | | | |
| 34 - 38 | | 0.08 | | | |
| 38 - 42 | | 0.04 | | | |
| TOTAL | | 1.00 | | | |

Hallar: $D_1$, $Q_3$, $P_{85}$

50. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 20 - 24 | 2 | | | | |
| 24 - 28 | 27 | | | | |
| 28 - 32 | 13 | | | | |
| 32 - 36 | 6 | | | | |
| 36 - 40 | 2 | | | | |
| TOTAL | | | | | |

Hallar: $R, S^2, S, CV, D_2, Q_3, P_{72}$

51. De dos bases de datos, se eligen en forma aleatoria 5 revistas científica, la variable en estudios es el número de artículos por revistas; los datos son:

$A = \{7, 2, 13, 3, 4\}$ , $B = \{11, 12, 11, 13, 11\}$,

¿Cual presenta mayor variabilidad?

52. La variancia de 10 datos es 5, si a cada dato se le incrementa en un 20% más 2 unidades. Hallar el nuevo valor de la variancia.

53. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18 - 22 | | 0.06 | | | |
| 22 - 26 | | 0.52 | | | |
| 26 - 30 | | 0.20 | | | |
| 30 - 34 | | 0.10 | | | |
| 34 - 38 | | 0.08 | | | |
| 38 - 42 | | 0.04 | | | |
| TOTAL | | 1.00 | | | |

Hallar: $D_4$, $Q_1$, $P_{87}$

54. La tabla contiene datos sobre las edades de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 20 - 24 | 2 | | | | |
| 24 - 28 | 6 | | | | |
| 28 - 32 | 13 | | | | |
| 32 - 36 | 28 | | | | |
| 36 - 40 | 1 | | | | |
| TOTAL | 50 | | | | |

Hallar: $R, S^2, S, CV, D_3, Q_1, P_{78}$

55. De dos bases de datos, se eligen en forma aleatoria 5 revistas científica, la variable en estudios es el número de artículos por revistas; los datos son:

$A = \{11, 12, 11, 13, 11\}$ , $B = \{7, 2, 13, 3, 4\}$,

¿Cual presenta mayor variabilidad?

56. La variancia de 10 datos es 5, si a cada dato se le incrementa en un 30% más 4 unidades. Hallar el nuevo valor de la variancia. (4PTOS)

57. Sea la tabla de datos: X=grado de instrucción del eje de familia; Y= número de hijos

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X | Sup | prim | prim | sec | sec | sup | sec | prim | Sup | sup |
| Y | 1 | 3 | 2 | 2 | 3 | 2 | 1 | 1 | 1 | 3 |

Elaborar la tabla de doble entrada

58. La media aritmética y variancia de 4 datos es 6 y $\frac{10}{3}$, respectivamente, una revisión de los datos, se encontró que en lugar de ingresar a la calculadora el valor 8 se ingreso por error el valor 4. Con esta información, calcular la media aritmética y variancia corregidas.

59. Una muestra tomada al azar de 50 revistas científicas en biología, proporciona los siguientes datos con respecto al número de artículos por revistas.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 12 | 12 | 13 | 13 | 13 | 13 | 13 |
| 13 | 13 | 13 | 13 | 13 | 13 | 13 | 14 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 14 | 14 | 14 | 15 | 15 | 15 | 15 | 15 |
| 15 | 15 | 15 | 15 | 15 | 15 | 15 | 16 |
| 16 | 16 | | | | | | |

Calcular:
La media aritmética, la mediana, la moda, el cuarto decil, el primer cuartil y el setenta y dos avo percentil del número de artículos por revistas.

60. La tabla contiene datos sobre la edad de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 18-22 | | 0.10 | | | |
| 22-26 | | “a” | | | |
| 26-30 | | 0.35 | | | |
| 30-34 | | “b” | | | |
| 34-38 | | 0.1 | | | |
| Total | | | | | |

Si la media aritmética es 28.2 años.
Calcular, la mediana, la moda, el 3er. decil, el 1er cuartil y el 87 avo percentil.

61. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio el número de artículos de 60 revistas científicas sobre biología elegidas al azar.

| X=N° de artículos por revistas | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 21 | 5 | | | | |
| 17 | 5 | | | | |
| 19 | 30 | | | | |
| 18 | 10 | | | | |
| 20 | 10 | | | | |
| Total | 60 | | | | |

Hallar: la media, mediana, moda, el 3er. decil, el 1er cuartil y el 87 avo percentil y el coeficiente de variación.

62. De una muestra de 80 datos el mayor y menor son 56.26 y 12.38. respectivamente. Determinar los intervalos de clases para la tabla de distribución de frecuencias.

63. Sean los datos:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 12 | 13 | 14 | 15 | 13 | 16 | 17 |
| 15 | 18 | 19 | 17 | | | | |

Hallar: la media, mediana, moda, el 3er. decil, el 1er cuartil y el 87 avo percentil y el coeficiente de variación.

64. Sean 50 datos, tal que $\bar{x} = 2$, además $\sum_{i=1}^{50} (x_i - \bar{x})^2 = 14.7$. Hallar su coeficiente de variación.

65. La tabla contiene datos sobre la edad de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| | | 0.10 | | | 22 |
| | 12 | | | | 26 |
| | | 0.20 | | | |
| | | 0.50 | | | |
| | 4 | | | | |
| Total | | | | | |

a. Completar la tabla

b. Hallar la media aritmética, mediana y moda

c. Que porcentajes de usuarios tienen edades entre 22 y los 30 años

d. Hallar la variancia y el coeficiente de variabilidad.

66. Sean los conjuntos de datos:

$$A = \{4, 12, 8, 7, 10, 8\}$$

$$B = \{3, 13, 9, 8, 10, 11\}$$

Indicar que conjunto de datos presenta mayor variabilidad o dispersión.

67. La tabla contiene datos sobre la edad de usuarios que concurren a un centro de información:

| Edad (años) | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 20-24 | | 0.10 | | | |
| 24-28 | | "a" | | | |
| 28-32 | | 0.30 | | | |
| 32-36 | | "b" | | | |
| 36-40 | | 0.10 | | | |
| Total | | | | | |

Si la media aritmética es 30 años.
Calcular, la mediana, la moda, el 3er decil, el 1er cuartil y el 85 avo percentil de las edades.

68. La media aritmética y variancia de 4 datos es 6 y $\frac{10}{3}$, respectivamente, una revisión de los datos, se encontró que en lugar de ingresar a la calculadora el valor 8 se ingreso por error el valor 4. Con esta información, calcular la media aritmética y variancia corregidas.

69. La siguiente tabla de distribución de frecuencias, tiene con variable en estudio el número de artículos de 80 revistas científicas sobre biología elegidas al azar.

| X=N° de artículos por revistas | $f_i$ | $h_i$ | $F_i$ | $H_i$ | $x_i'$ |
|---|---|---|---|---|---|
| 21 | 5 | | | | |
| 17 | 5 | | | | |
| 19 | 20 | | | | |
| 18 | 40 | | | | |
| 20 | 10 | | | | |
| Total | 80 | | | | |

Hallar: la media, mediana, moda, la variancia y el coeficiente de variación.

# CAPÍTULO 5
# ANÁLISIS EXPLORATORIO DE DATOS

## INTRODUCCIÓN

Esta técnica trata de la descripción de las características de la variable en estudio a través de sus datos, con el objetivo de conocer la dispersión, simetría o asimetría, ubicación de la mediana, así como la presencia de datos discordante (autlier). Para su análisis se presenta el diagrama de tallos y hojas, y el diagrama de cajas.

## 5.1. GRÁFICO DE TALLOS Y HOJAS

Esta técnica de análisis de datos, tiene el mismo propósito del histograma de frecuencias. Sin embargo una ventaja sobre el histograma es que conserva los datos en su forma original, ya que en la construcción del histograma de frecuencias, los datos originales ya no son considerados sino solo las frecuencias absolutas en las clases respectivas. Esta técnica de tallos y hojas proporciona información respecto: a la Simetría o Asimetría de un conjunto de datos, su dispersión o concentración y la presencia o ausencia de datos extremos o anormales (datos muy bajos o muy altos de lo común).

## 5.2. CONSTRUCCIÓN DEL GRÁFICO DE TALLOS Y HOJAS

Los datos de la muestra se ordenan en forma ascendente, luego un dato se divide en dos partes, tallos y hojas; los primeros dígitos del dato como un tallo y el último digito como una hoja de dicho tallo, las hojas pueden tomar valores entre 0 y 9. Luego los valores del tallos se colocan de menor a mayor en forma vertical y sus valores correspondientes a las hojas, separados por un segmento de recta vertical, y en la parte inferior del GRÁFICO de tallos y hojas se coloca la unidad que utiliza la hoja

Ejemplos: descomponer los datos que a continuación se indica en tallo y hojas y colocar la unidad de la hoja:

$42 \rightarrow 4 / 2$ (El tallo es 4, de las decenas y la hoja 2 de la unidad)

$4.2 \rightarrow 4/5$(El tallo es 4 de la unidad y la hoja 2 de la décima)

$42.5 \rightarrow 42/5$ (El tallo es 42 de la unidad y la hoja 5 de la décima)

$42.35 \rightarrow 423/5$(El tallo es 423 de la décima y la hoja 5 de la centésima)

$0.42 \rightarrow 4/2$ (El tallo es 4 de la décima y la hoja es 2 de la centésima)

$0.042 \rightarrow 4/2$ (El tallo es 4 de la centésima y la hoja es 2 de la milésima)

$0.0042 \rightarrow 4/2$ (El tallo es 4 de la milésima y la hoja es 2 de la diez milésima)

Ejemplo: Sean los datos de pesos de recién nacidos (kg)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1.5 | 1.6 | 1.9 | 2.0 | 2.1 | 2.1 | 2.3 | 2.4 |
| 2.5 | 2.5 | 2.6 | 2.7 | 2.8 | 2.9 | 3.0 | 3.1 |
| 3.1 | 3.2 | 3.2 | 3.2 | 3.2 | 3.3 | 3.3 | 3.3 |
| 3.4 | 3.4 | 3.5 | 3.6 | 3.6 | 3.7 | 3.7 | 3.9 |
| 4.1 | 4.1 | | | | | | |

Elaborar el diagrama de tallos y hojas, comentarios

| | |
|---|---|
| 1 | 569 |
| 2 | 01134556789 |
| 3 | 011222233344566779 |
| 4 | 11 |

Siendo la unidad de hoja=0.1 (decima), el tallo es la unidad

Interpretación de gráfico de tallos y hoja

En este caso la distribución de los datos es asimétrica a la izquierda o asimetría negativa existe poca variabilidad y no hay datos extremos.

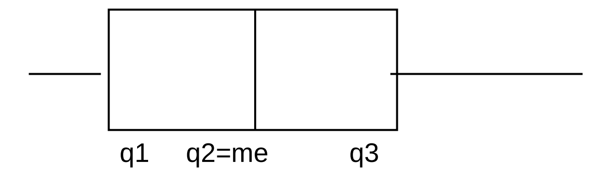
En este caso los tallos lo conforman los primeros dígitos y las hojas los últimos dígitos. El uso del GRÁFICO de tallos y hojas es igual al del histograma de frecuencias, la única diferencia esta en que del GRÁFICO de tallos y hojas se pueden recuperar los datos, lo cual se logra al juntar los tallos con su respectiva hoja y multiplicándolo por la unidad de la hoja.

## 5.3. DIAGRAMA DE CAJA

El diagrama de caja al igual que el histograma de frecuencias y el gráfico de tallos y hojas, permite tener una idea visual de la distribución

de los datos; es decir, si hay simetría, ver el grado de variabilidad existente y por ultimo identificar la presencia de datos extremos (datos anormales). El diagrama de de caja es útil para comparar grupos de datos, es una alternativa Gráfica a la prueba de t-Student ( si se compara dos grupos de datos) o a la prueba de F-Snedecor del análisis de variancia, si se compara mas de dos grupos de datos, lo anterior es posible debido a que se pueden hacer múltiples diagramas de cajas en una misma Gráfica, en cambio los diagramas de tallos y hojas salen en secuencia uno por paginas, esto es usando los paquetes estadísticos MINITAB o SPSS.

Características de diagrama de caja




1. La línea central de la caja representa el valor de la mediana (q2=me)
2. Los lados de la caja representa el 1er. Cuartel (q1) y tercer cuartel (q3).
3. Si el valor de la mediana se encuentra en el centro, entonces los datos se distribuyen simétricamente.
4. Si el valor de la mediana tiende a q1, existe asimétrica hacia la derecha (asimetría positiva), si el valor de la mediana tiende a q3, existe asimétrica hacia la izquierda ( asimetría negativa)
5. Si la caja no es alargada, nos indica que no hay mucha variabilidad en los datos.
6. Si no hay datos extremos o anormales (datos muy altos o datos muy bajos de lo normal) , las líneas laterales de la caja llegan hasta el dato menor y el dato mayor.
7. Si hay datos extremos estos aparecen identificados en el gráfico y las líneas laterales llegan hasta los valores adyacentes a las fronteras interiores.
8. Las fronteras interiores se calculan por $Q_1 - 1.5RIQ \;\; y \; Q_1 + 1.5RIQ$, respectivamente, siendo $RIQ = Q_3 - Q_1$ el rango intercuartilico, las fronteras exteriores se calcula por $Q_1 - 3RIQ \;\; y \; Q_1 + 3RIQ$ .

9. Si un dato cae fuera de $Q_1 - 3RIQ$ y $Q_1 + 3RIQ$ , se dice que es un dato extremo, en caso contrario es un dato extremo moderado. Un dato extremo se representa por "0" y un dato moderado por " $x$"

Ejemplos

Sean los datos de tres variables

| | X | Y | Z |
|---|---|---|---|
| 1 | 1.2 | 1.1 | 1.0 |
| 2 | 1.3 | 2.0 | 1.1 |
| 3 | 2.0 | 2.1 | 2.0 |
| 4 | 2.0 | 3.0 | 2.0 |
| 5 | 2.0 | 3.0 | 2.1 |
| 6 | 2.0 | 3.1 | 3.0 |
| 7 | 2.1 | 4.0 | 3.0 |
| 8 | 2.1 | 4.0 | 3.1 |
| 9 | 3.0 | 4.1 | 3.1 |
| 10 | 3.0 | 4.1 | 3.1 |
| 11 | 3.1 | 4.2 | 3.2 |
| 12 | 3.1 | 4.2 | 4.0 |
| 13 | 4.0 | 4.2 | 4.0 |
| 14 | 4.1 | 5.0 | 4.1 |
| 15 | 5.0 | 5.0 | 5.0 |
| 16 | | 5.1 | 5.1 |
| 17 | | 5.1 | |

Representar los datos de las columnas en gráficos de tallos y hojas

Para la primera variable

| | |
|---|---|
| 2 | 1 23 |
| (6) | 2 000011 |
| 7 | 3 0011 |
| 3 | 4 01 |
| 1 | 5 0 |

Unidad de hoja=0.1

Para la Segunda columna:

| | |
|---|---|
| 1 | 1 1 |
| 3 | 2 01 |
| 6 | 3 001 |
| (7) | 4 0011222 |
| 4 | 5 0011 |

Unidad de hoja =0.1

Para la tercera columna:

| | |
|---|---|
| 2 | 1 01 |
| 5 | 2 001 |
| (6) | 3 001112 |
| 5 | 4 001 |
| 2 | 5 01 |

Unidad de hoja=0.1

Calcular de la media, mediana, desviación estándar para las tres columnas, usando las formulas ya conocidas:

| Columnas | Numero de datos | media | mediana | Desviacion estandar |
|---|---|---|---|---|
| 1 | 15 | 2.667 | 2.1 | 1.079 |
| 2 | 17 | 3.724 | 4.1 | 1.175 |
| 3 | 16 | 3.056 | 3.1 | 1.213 |

## 5.4. OTRAS MEDIDAS DE POSICIÓN

- PERCENTILES PARA DATOS SIN AGRUPAR

Son valores que divide al conjunto de datos ordenados ascendente o descendente en 100 partes iguales, la formula para calcular los percentiles es:

$$P_j = X_{(n+1)\frac{j}{100}} = X_i + a(X_{i+1} - X_i) \qquad (\alpha)$$

$$J = 1, 2, 3, 4, \ldots, 99$$

Siendo:

$P_j$ = Valor que divide a un conjunto de “n” datos ordenados, de tal forma que el $j\%$ de los datos son menores o iguales que el valor del percentil $P_j$ y $(100 - j)\%$ de los datos son mayores al valor de percentil $P_j$

$X_i$ = Valor del dato en la posición del entero del número $(n+1)\frac{j}{100}$

“a” =decimal del numero $(n+1)\frac{j}{100}$

- <u>CUARTILES $Q_j$ PARA DATOS SIN AGRUPAR</u>

Son valores que divide al conjunto de datos ordenados ascendente o descendente en 4 partes iguales, en la formula es:

$$Q_j = X_{(n+1)\frac{j}{4}} = X_i + a(X_{i+1} - X_i) \qquad J = 1, 2, 3$$

- <u>DECILES $(D_j)$ PARA DATOS SIN AGRUPAR</u>

Son valores que divide al conjunto de datos ordenados ascendente o descendente en 10 partes iguales, en la formula ( $\alpha$ ), $j$ toma los siguientes valores:

$$D_j = X_{(n+1)\frac{j}{10}} = X_i + a(X_{i+1} - X_i) \qquad J = 1, 2, 3,4,5,6,7,8,9$$

Ejemplo:

La variable peso (kg) de los recién nacidos

| | |
|---|---|
| 1 | 569 |
| 2 | 01134556789 |
| 3 | 011222233344566779 |
| 4 | 11 |

Siendo la unidad de hoja=0.1 n=34

Calcular:

Los cuartiles, el tercer decil y el percentil 72 avo, el rango intercuartilico, media, moda, desviación estandar

Utilizado las formula correspondiente, se tiene:

$$Q_1 = X_{0.25*35} = X_{8.75} = x_8 + 0.75(x_9 - x_8) = 2.4 + 0.75(2.5 - 2.4) = 2.475$$

$$Q_2 = X_{0.50*35} = X_{17.5} = x_{17} + 0.5(x_{18} - x_{17}) = 3.1 + 0.5(3.2 - 3.1) = 3.150$$

$$Q_3 = X_{0.75*35} = X_{26.25} = x_{26} + 0.25(x_{27} - x_{26}) = 3.4 + 0.25(3.5 - 3.4 = 3.425$$

Tercer cuartil

$$D_3 = X_{0.3*35} = X_{10.5} = x_{10} + 0.5(x_{11} - x_{10}) = 2.5 + 0.5(2.6 - 2.5) = 2.55$$

Setenta y dos percentil

$$P_{72} = X_{0.72*35} = X_{25.2} = x_{25} + 0.2(x_{26} - x_{25}) = 3.4 + 0.2(3.4 - 3.4) = 3.4$$

$$RIQ = Q_3 - Q_1 = 3.425 - 2.475 = 0.95$$

$$\bar{x} = 2.965\ kg$$

$$m_o = 3.2\ kg$$

$$s = 0.687\ kg$$

$$cv = \frac{0.687}{2.965} * 100\% = 23.17\%$$

- CÁLCULO DE PERCENTILES $P_j$ PARA DATOS AGRUPADOS

Para datos discretos

Ejemplo

| X=No de artículos por revistas | $f_i$ | $F_i$ | |
|---|---|---|---|
| 14 | 10 | 10 | $x_1 \rightarrow x_{10} = 14$ |
| 15 | 15 | 25 | $x_{11} \rightarrow x_{25} = 15$ |
| 16 | 20 | 45 | $x_{25} \rightarrow x_{45} = 16$ |
| 17 | 18 | 63 | $x_{46} \rightarrow x_{63} = 17$ |
| 18 | 14 | 77 | $x_{64} \rightarrow x_{77} = 18$ |
| 19 | 6 | 83 | $x_{78} \rightarrow x_{83} = 19$ |
| total | 83 | | |

Calcular: $D_7, \quad Q_3, \quad P_{72}$

$$D_7 = X_{(83+1)\frac{7}{10}} = X_{58.8} = X_{58} + 0.8(X_{59} - X_{58}) = 17 + 0.8(17 - 17) = 17$$

El 70% de las revistas, tienen menor o igual a 17 artículos, el 30% de las revistas tienen mayor a 17 artículos.

$$Q_3 = X_{(83+1)\frac{3}{4}} = X_{63} = 17$$

El 75% de las revistas, tienen menor o igual a 17 artículos, el 25% de las revistas tienen mayor a 17 artículos.

$$P_{72} = X_{(83+1)\frac{72}{100}} = X_{60.48} = X_{60} + 0.48(X_{61} - X_{60}) = 17 + 0.48(17 - 17) = 17$$

El 72% de las revistas, tienen menor o igual a 17 artículos, el 28% de las revistas tienen mayor a 17 artículos

<u>Para datos continuos</u>

La formula es:

$$P_j = L_i + \left(\frac{\frac{nj}{100} - F_{i-1}}{f_i}\right) * A$$

$$J = 1, 2, 3, 4, \ldots, 99$$

Dónde:

$L_i =$ Límite Inferior de la clase que contiene al valor del percentil

$P_j =$ Valor del percentil que acumular el $j\%$ de los datos de la muestra

$A =$ Ancho de clase

$F_{i-1} =$ Frecuencia acumulada absoluta de la clase anterior a la clase que contiene al valor del percentil

$f_i =$ Frecuencia absoluta de la clase que contiene al valor del percentil

Nota:

La primera frecuencia absoluta acumulada o frecuencia relativa acumulada que sea superior al valor $\frac{nj}{100}$ o $\frac{j}{100}$, será la clase que contiene al valor del percentil $P_j$.

Es decir:

$$F_i \geq \frac{nj}{100} \quad o \quad H_i \geq \frac{j}{100}$$

Para los Cuartiles será:

$$Q_j = L_i + \left(\frac{\frac{nj}{4} - F_{i-1}}{f_i}\right) * A$$

$$J = 1, \quad 2, \quad 3$$

La relación para ubicar en que clase se encuentra el valor del cuartil es:

$$F_i \geq \frac{nj}{4} \quad o \quad H_i \geq \frac{j}{4}$$

Para los Deciles será:

$$D_j = L_i + \left(\frac{\frac{nj}{10} - F_{i-1}}{f_i}\right) * A$$

$J = 1, 2, 3, 4, 5, 6, 7, 8, 9$

La relación para ubicar en que clase se encuentra el valor del decil es:

$$F_i \geq \frac{nj}{10} \quad o \quad H_i \geq \frac{j}{10}$$

Ejemplo aplicativo

La tabla de distribución de frecuencias contiene a las edades (años) de los usuarios que concurren a un centro de información: Calcular el 1er cuartil, 3er cuartil; 4to. decil y el 65avo percentil.

| Edades de usuarios | $f_i$ | $F_i$ |
|---|---|---|
| 18.5-22.5 | 5 | 5 |
| 22.5-26.5 | 9 | 14 |
| 26.5-30.5 | 12 | 26 |
| 30.5-34.5 | 20 | 46 |
| 34.5-38.5 | 13 | 59 |
| 38.5-42.5 | 10 | 69 |
| 42.5-47.5 | 4 | 73 |
| Total | 73 | |

Cálculo del 1er cuartil

$$\frac{nj}{4} = \frac{73*1}{4} = 18.25 \quad \rightarrow F_i \geq 18.25 \rightarrow i = 3 \quad clase$$

$$Q_1 = L_3 + \left(\frac{\frac{nj}{4} - F_{3-1}}{f_3}\right) * A = 26.5 + \left(\frac{18.25 - 14}{12}\right) * 4 = 27.92 \ años$$

El 25% de los usuarios tienen edades menores o iguales a 27.92 años, y el otro 75% de los usuarios tienen edades superior a los 27.92 años.

Cálculo del 3er cuartil

$$\frac{nj}{4} = \frac{73*3}{4} = 54.75 \quad \rightarrow F_i \geq 54.75 \rightarrow i = 5 \quad clase$$

$$Q_3 = L_5 + \left(\frac{\frac{nj}{4} - F_{5-1}}{f_5}\right) * A = 34.5 + \left(\frac{54.75 - 46}{13}\right) * 4 = 37.19 \text{ años}$$

Interpretación

El 75% de los usuarios tienen edades menores o iguales a 37.19 años y el otro 25% de los usuarios tienen edades superiores a 37.192 años.

Cálculo del 4to. decil

$$\frac{nj}{10} = \frac{73*4}{10} = 29.2 \quad \rightarrow F_i \geq 29.2 \rightarrow i = 4$$

$$D_4 = L_4 + \left(\frac{\frac{nj}{4} - F_{4-1}}{f_4}\right) * A = 30.5 + \left(\frac{29.2 - 26}{20}\right) * 4 = 31.14 \text{ años}$$

Interpretación

El 40% de los usuarios tienen edades menores o iguales a 31.14 años y el otro 60% de los usuarios tienen edades superiores a 31.14 años.

Cálculo del 65 avo percentil

$$\frac{nj}{100} = \frac{73*65}{100} = 47.45 \quad \rightarrow F_i \geq 47.45 \rightarrow i = 5$$

$$P_{65} = L_5 + \left(\frac{\frac{nj}{100} - F_{5-1}}{f_5}\right) * A = 34.5 + \left(\frac{47.45 - 46}{13}\right) * 4 = 34.95 \text{ años}$$

Interpretación

El 65% de los usuarios tienen edades menores o iguales a 34.95 años y el otro 35% de los usuarios tienen edades superiores a 34.95 años.

## 5.5. OTRA MEDIDA DE VARIABILIDAD

### RANGO INTERCUARTIL

El rango intercuartilico para un conjunto de datos, es la diferencia entre el tercer y el primer cuartil, y su valor expresa la variabilidad en el 50% central de los datos.

$$RIC = Q_3 - Q_1$$

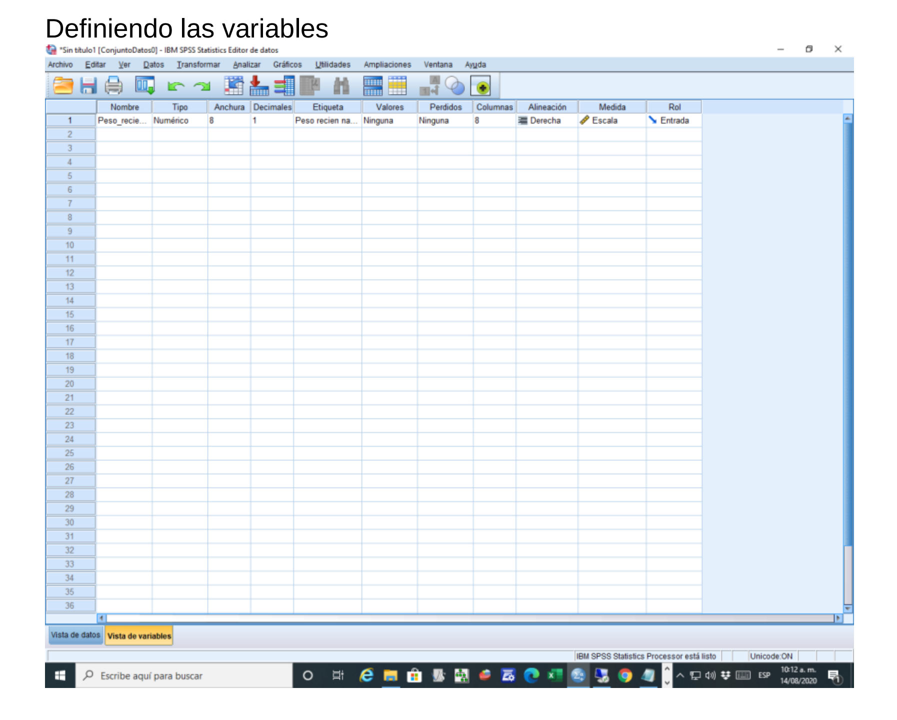
Ejemplo

De la tabla anterior el rango intercuartilico es:

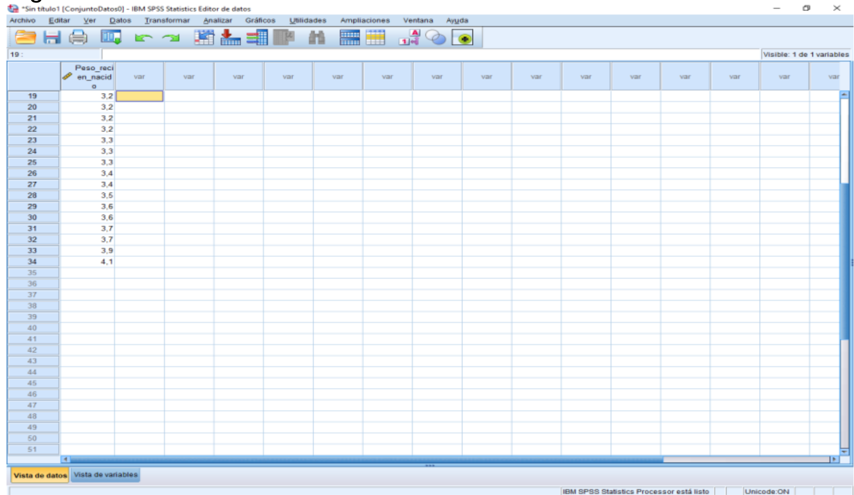
$RIC = Q_3 - Q_1 = 37.19 - 27.92 = 9.27$ Años es la variabilidad de las edades para el 50% central de los datos

Aplicando el SPSS para realizar el gráfico de tallo y hojas y el diagrama de cajas

Definiendo las variables



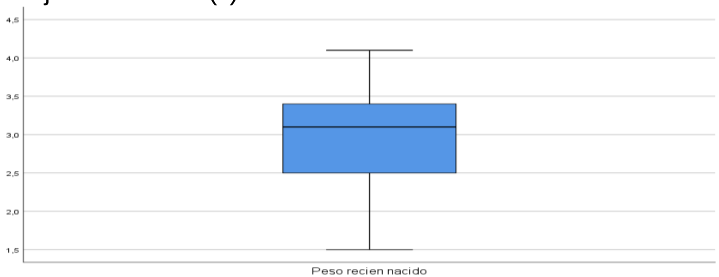

Ingresando los datos




En el Menú del SPSS escoger ANALIZAR-ESTADISTICOS DESCRIPTIVOS—EXPLORAR-pasar la variable Peso_recien_nacido a la casilla de la derecha-GRÁFICO-marcar de tallo y -CONTINUAR-ACEPTAR

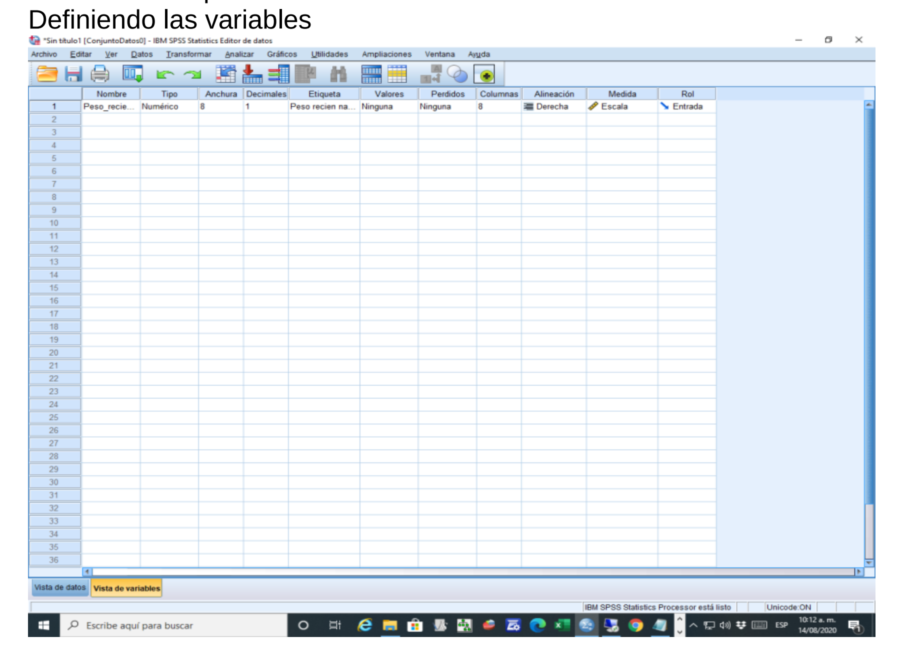
Peso recién nacido Gráfico de tallo y hojas

```
Frecuencia Stem &  Hoja
     .00        1 .
    3.00        1 .  569
    5.00        2 .  01134
    7.00        2 . 5566789
   12.00        3 . 011222233344
    6.00        3 . 566779
    1.00        4 .  1
 Ancho del tallo:       1.0
 Cada hoja:        1 caso(s)
```

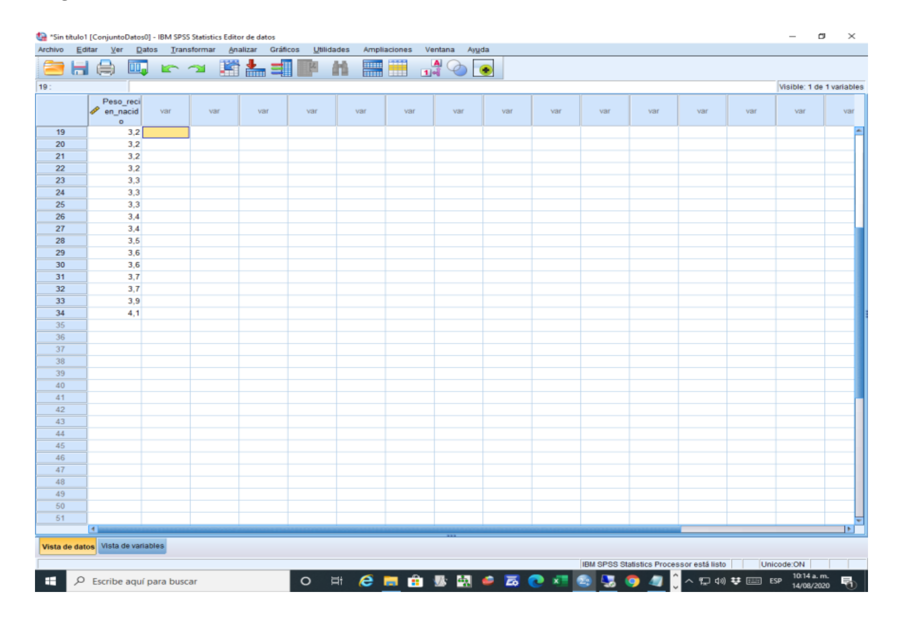


## Cálculo de los percentiles con SPSS

## Definiendo las variables




## Ingresando los datos



Se elabora la tabla de frecuencias para datos agrupados:

10. Cálculo del Rango” R “.

R = peso mayor –peso menor.

R = 4.1-1.5 = 2.6

11. Cálculo del número de clases (K).

$$K = 1 + 3.3Log(34) = 6.05388$$

Como el valor que toma K debe ser un número entero, usar el redondeo simple (redondeo por defecto), se tiene que K = 6.

OBSERVACIÓN

El redondeo simple significa, que si un número tiene como primer decimal un número mayor o igual a 5, entonces la cifra entera del número se incrementa en una unidad.

Ejemplo:

Si, K = 6.7, usando el redondeo simple K = 7;
Si, K = 6.4, usando el redondeo simple K = 6;

12. Cálculo del ancho del intervalo de clase (A)

El ancho del intervalo de clase se obtiene a través de la siguiente fórmula:

$$A = \frac{R}{K}$$

Para los datos se tiene:

$$A = \frac{R}{K} = \frac{2.6}{6} = 0.43333$$

El valor que toma “A”, debe tener el mismo número de decimales de los datos originales. Por redondeo simple se tiene:

$$A = 0.4$$

$$E = AK - R = 0.4(6) - 2.6 = -0.2$$

Como el error es negativo, se debe incrementar en una clase más(k=7)

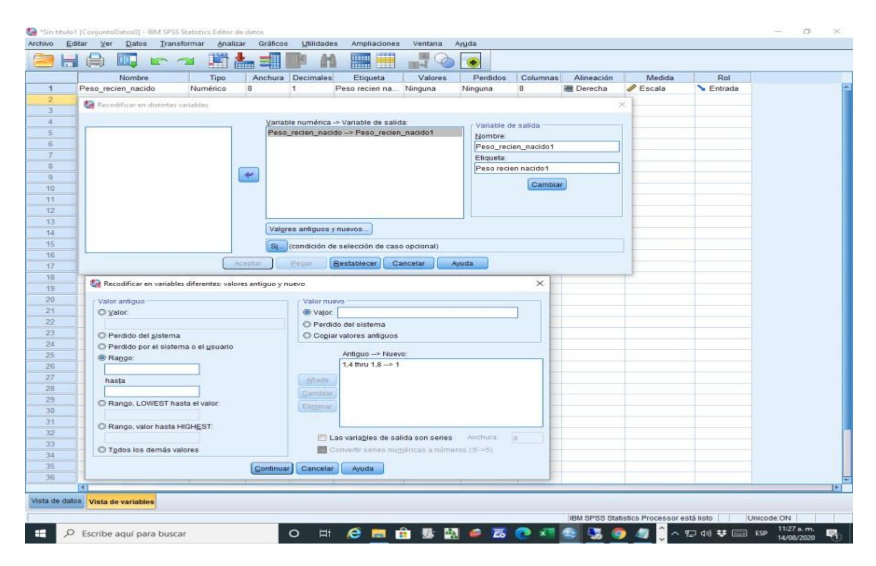
$$A = \frac{2.6}{7} = 0.4$$

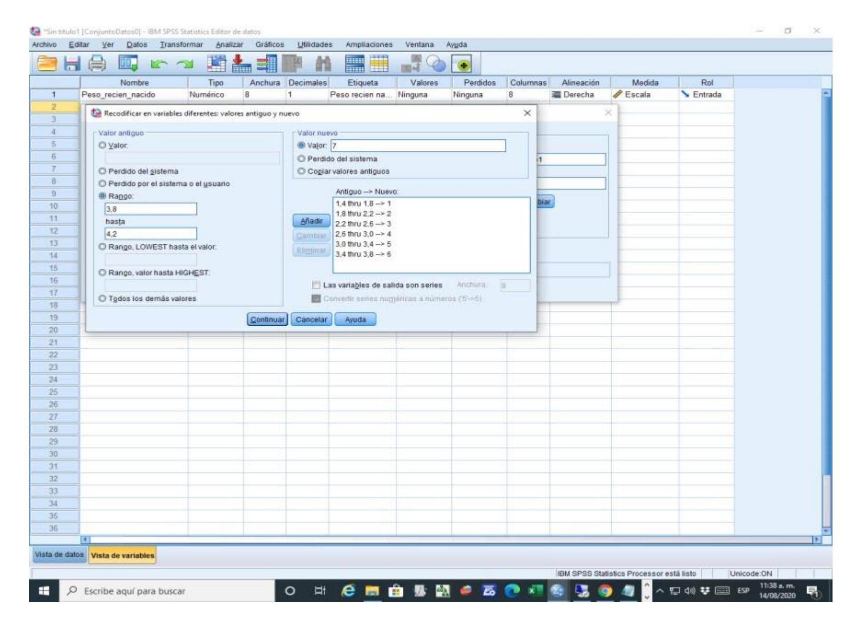
$$E = AK - R = 0.4(7) - 2.6 = 0.2$$

Cuando el error es positivo, la tabla tendrá 7 clases y ancho 0.4

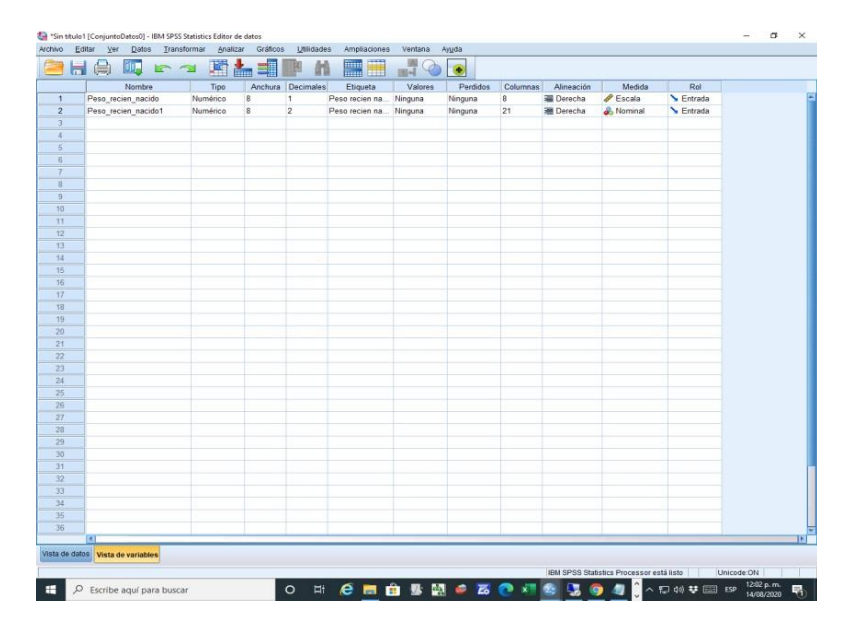
$$valor\ de\ arranque = dato\ menor - \frac{E}{2} = 1.5 - \frac{0.2}{2} = 1.4$$

En el menú del SPSS, escoger la opción TRANSFORMAR-RECODIFICAR EN DISTINTAS VARIABLES-pasar la variable peso-recién-nacido al casillero de la derecha, en la derecha en donde indica variable de salida Nombre: poner peso_recien_nacido1 y en la etiqueta: poner Peso recien nacido1 – PRESIONAL LA OPCION CAMBIAR-VALORES ANTIGUOS Y NUEVOS-RANGO-poner desde 1,4 hasta 1,8, luego en VALOR poner 1 y AÑADIR, así sucesivamente continua con todos los rangos de 3,8 hasta 4,2.

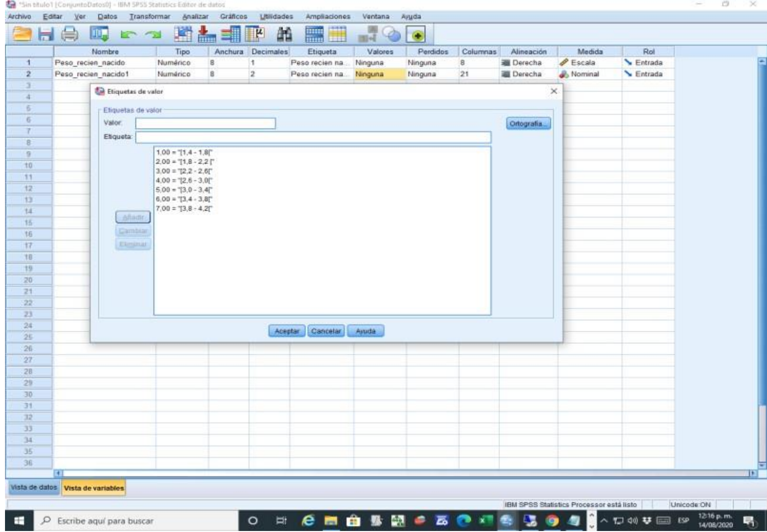





Cuando se termine de ingresar los rangos presionar continuar y aceptar.
Y se crea una variable Peso_recien_nacido_1.

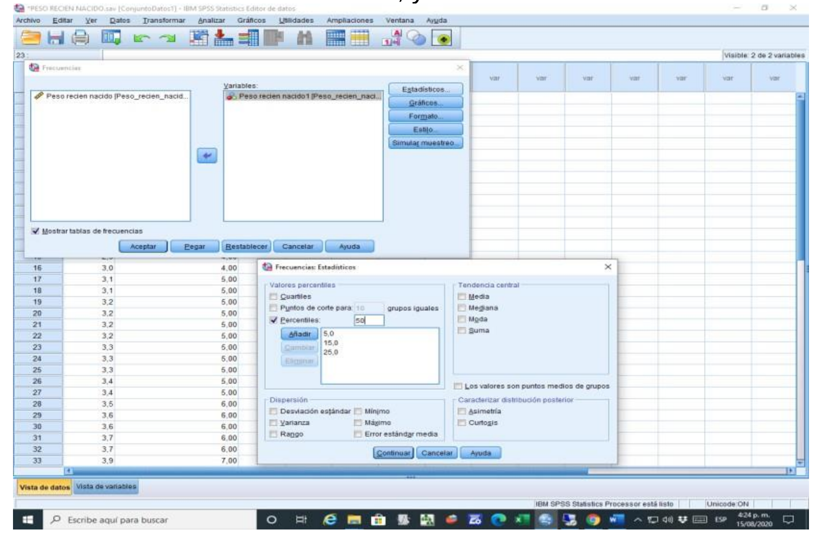


En la variable Peso_recien_nacido_1, presionar la opción valores e ingresar:




Y Poner aceptar.

En el Menú del SPSS escoger la opción ANALIZAR-ESTADISTICOS DESCRIPTIVOS-FRECUENCIA-pasar la variable Peso recien nacido 1 al casillero derecho-ESTADISTICOCO-seleccionar percentiles-en casillero de la derecha indicar que percentil se quiere calcular y poner añadir-CONTINUAR- ACEPTAR, y nos muestra los resultados



**Estadísticos**

Peso recien nacido1

| | | |
|---|---|---|
| N | Válido | 34 |
| | Perdidos | 0 |
| Percentiles | 5 | 1,0000 |
| | 15 | 2,0000 |
| | 25 | 3,0000 |
| | 50 | 5,0000 |

5.6. Ejercicios propuestos.

1. Los siguientes datos corresponden a las edades (años) de 50 pacientes elegidos al azar de un centro de salud de un mes determinados

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 33,7 | 27,6 | 34,1 | 24,6 | 25,2 | 32,3 | 22,8 | 32,4 | 24,5 | 26,9 | 25,3 |
| 28,0 | 16,4 | 21,0 | 23,9 | 24,4 | 24,2 | 18,2 | 29,3 | 22,7 | 33,5 | 24,6 |
| 21,0 | 17,7 | 23,8 | 24,8 | 32,6 | 40,2 | 13,7 | 28,7 | 31,6 | 21,8 | 22,6 |
| 25,7 | 27,0 | 21,6 | 23,2 | 28,2 | 26,9 | 24,1 | 28,3 | 39,1 | 25,5 | 24,3 |
| 15,0 | 19,3 | 30,5 | 25,1 | 26,1 | 27,3 | | | | | |

Elaborar el gráfico de tallos-hojas y caja, y calcular la media, mediana, moda, los cuarties, 4to y 7mo avo deciles y el 72 avo percentil, rango intercuartilico. Realice comentarios de sus resultados.

2. Los siguientes datos corresponden a las estaturas (metros) de 50 alumnos elegidos al azar de un instituto de estudios superior.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1,52 | 1,64 | 1,69 | 1,48 | 1,53 | 1,56 | 1,86 | 1,59 | 1,71 | 1,76 | 1,56 |
| 1,47 | 1,56 | 1,72 | 1,60 | 1,80 | 1,68 | 1,70 | 1,46 | 1,50 | 1,69 | 1,75 |
| 1,71 | 1,76 | 1,54 | 1,73 | 1,58 | 1,67 | 1,80 | 1,75 | 1,68 | 1,67 | 1,49 |
| 1,44 | 1,68 | 1,76 | 1,66 | 1,42 | 1,53 | 1,69 | 1,72 | 1,73 | 1,74 | 1,71 |
| 1,67 | 1,69 | 1,67 | 1,54 | 1,63 | 1,65 | | | | | |

Elaborar el gráfico de tallos-hojas y caja, y calcular la media, mediana, moda, los cuarties, 3er y 6to deciles y 54 avo percentil, rango intercuartilico. Realice comentarios de sus resultados.

3. Los siguientes datos corresponden a los montos por ventas semanal (miles de nuevos soles) de 50 grifos elegidos al azar de cierta ciudad.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 56,0 | 65,6 | 50,0 | 61,3 | 78,5 | 66,0 | 60,0 | 58,0 | 48,2 | 55,4 | 43,0 |
| 45,3 | 50,0 | 56,5 | 56,1 | 75,9 | 69,2 | 51,9 | 53,3 | 74,0 | 41,0 | 55,1 |
| 50,2 | 50,9 | 62,0 | 53,7 | 43,4 | 69,6 | 75,5 | 51,9 | 45,0 | 74,6 | 56,6 |
| 63,2 | 66,4 | 48,8 | 65,8 | 65,0 | 49,8 | 55,2 | 65,9 | 60,8 | 42,0 | 74,6 |
| 73,6 | 43,9 | 40,2 | 40,9 | 41,8 | 54,5 | | | | | |

Elaborar el gráfico de tallos-hojas y caja, y calcular la media, mediana, moda, los cuarties, 4to y 7mo deciles y el 35 avo y 84 avo percentiles, rango intercuartilico. Realice comentarios de sus resultados.

4. Los siguientes datos corresponden al número de empleados que laboran administrativamente en 50 garitas de peaje elegidos al azar de cierta región de un país.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 4 | 5 | 6 | 5 | 6 | 7 | 6 | 7 | 8 | 7 | 8 9 5 6 7 6 5 7 |
| 6 | 6 | 5 | 6 | 7 | 6 | 7 | 8 | 7 | 8 | 9 | 5 6 7 6 5 7 6 7 |
| 6 | 7 | 8 | 7 | 8 | 9 | 5 | 6 | 7 | 6 | 5 | 7 |

Elaborar el gráfico de tallos-hojas y caja, y calcular la media, mediana, moda, los cuartiles, 3er y 8vo deciles y 36 avo y 73 avo percentiles, rango intercuartilico. Realice comentarios de sus resultados.

# CAPÍTULO 6
# COEFICIENTES DE ASIMETRÍA Y CURTOSIS

## INTRODUCCIÓN

En esta parte se estudiará la forma de una distribución de frecuencias. Las medidas de las formas de una distribución se clasifica en dos: medidas de asimetría (desplazamiento de la curva hacia la derecha o izquierda) y medidas de curtosis (concentración de los datos con respecto a la media aritmética).

## 6.1. CONCEPTO DE SIMETRÍA

Cuando al trazar una recta vertical (en el diagrama de líneas o histograma de frecuencias de una variable, según sea esta discreta o continua), por el valor de la media aritmética, esta vertical se le llama eje de simetría; decimos que la distribución es simétrica cuando a ambos lados de la media aritmética existe aproximadamente el mismo número de valores de la variable, equidistantes de dicha media dos a dos; es decir que cada par de valores equidistantes tiene aproximadamente la misma frecuencia absoluta. En caso contrario, dicha distribución será asimétrica o diremos que presenta asimetría.

## 6.2. MEDIDAS DE ASIMETRÍA

Las medidas de asimetría indican la deformación horizontal de la curva. Cuando una distribución de frecuencias esta inclinada o alargada hacia la derecha, se llama asimetría a la derecha o positiva, y si esta alargada o inclinada hacia el lado izquierdo, se llama asimetría a la izquierda o negativa. Algunos criterios para determinar la asimetría de una distribución son:

- COEFICIENTE DE ASIMETRÍA DE FISHER

Este valor nos indica el desplazamiento de la distribución de los datos con respecto a la media aritmética y se calcula usando las siguientes fórmulas:

Para datos sin  agrupador:

$$A_f = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{ns^3}$$

Para datos discretos agrupados

$$A_f = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^3 f_i}{ns^3}$$

Para datos continuos agrupados

$$A_f = \frac{\sum_{i=1}^{k} (x'_i - \bar{x})^3 f_i}{ns^3}$$

Según el valor de $A_f$, los datos tendrá una distribución asimétrica a la izquierda, asimetría a la derecha o será simétrica, esto es:

$A_f \approx 0$ , la distribución de los datos será simétrica. Los datos se desplazan a la izquierda y derecha de la media aritmética.

$A_f > 0$ , la distribución de los datos será asimétrica a la derecha o asimetría positiva. Los datos se desplazan a la derecha de la media aritmética.

$A_f < 0$ , la distribución de los datos será asimétrica a la izquierda o asimetría negativa. Los datos se desplazan a la izquierda de la media aritmética.

Ejemplo

La tabla de distribución de frecuencia, corresponde a edades (años) de usuarios que concurren a un centro de información.

| Edades (años) | $f_i$ | $x_i'$ | $F_i$ |
|---|---|---|---|
| 18.5-22.5 | 3 | 20.5 | 3 |
| 22.5-26.5 | 4 | 24.5 | 7 |
| 26.5-30.5 | 14 | 28.5 | 21 |
| 30.5-34.5 | 12 | 32.5 | 33 |
| 34.5-38.5 | 5 | 36.5 | 38 |
| 38.5-42.5 | 4 | 40.5 | 42 |
| Total | 42 | | |

La distribución de frecuencias de las edades, que tipo de asimetría tiene.

Cálculos:

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i' f_i}{n} = 30.7857143$$

$$s^2 = \frac{\sum_{i=1}^{6} (x_i')^2 f_i - n(\bar{x})^2}{n-1} = 27.4285714 \rightarrow s = \sqrt{27.4285714} = 5.23722937$$

$$\sum_{i=1}^{6} (x_i' - \bar{x})^3 f_i = 235.101993$$

$$ns^3 = 42 * (5.23722937)^3 = 6033.28824$$

$$A_f = \frac{235.101993}{6033.28824} = 0.0389 > 0$$

La distribución de las edades es ligeramente asimétrica positiva, es decir los datos tienen un desplazamiento casi a ambos lados de la media aritmética.

- COEFICIENTE DE ASIMETRÍA DE PEARSON

Este coeficiente de asimetría se utiliza cuando se conoce los valores de la media aritmética, mediana, moda y desviación estándar.

$$A_p = \frac{\bar{x} - m_o}{s} = \frac{3(\bar{x} - m_e)}{s}$$

Esta formula se aplica, cuando la distribución de frecuencias de los datos unimodal.

$A_p \approx 0$ , la distribución de los datos será simétrica. Los datos se desplazan a la izquierda y derecha de la media aritmética.

$A_p > 0$ , la distribución de los datos será asimétrica a la derecha o asimetría positiva. Los datos se desplazan a la derecha.

$A_p < 0$ , la distribución de los datos será asimétrica a la izquierda o asimetría negativa. Los datos se desplazan a la izquierda.

De la tabla de distribución de frecuencias anterior, calcular el coeficiente de asimetría de Pearson.

De la tabla:

$$\bar{x} = 30.7857143$$

$$m_e = 30.5$$

$$s = 5.23722937$$

$$A_P = \frac{3(30.7857143 - 30.5)}{5.23722937} = 0.1636$$

Las edades, tienen una distribución asimétrica a la derecha.

- <u>COEFICIENTE DE ASIMETRÍA DE BOWLEY</u>

Este coeficiente se utiliza, cuando no es posible calcular la media y desviación estándar

$$A_b = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

Siendo:

$Q_1 =$ Valor del primer cuartil

$Q_2 =$ Valor del tercer cuartil

$Q_3 =$ Valor del segundo cuartil

$A_b \approx 0$ , la distribución de los datos será simétrica

$A_b > 0$ , la distribución de los datos será asimétrica a la derecha o asimetría positiva.

$A_b < 0$ , la distribución de los datos será asimétrica a la izquierda o asimetría negativa.

De la tabla de distribución de frecuencias anterior, calcular el coeficiente de asimetría de Bowley.

De la tabla: $Q_1 = 27.5 \quad Q_2 = 30.5 \quad Q_3 = 34$ , reemplazando valores

$$A_b = \frac{27.5 + 34 - 2(30.5)}{34 - 27.5} = 0.08$$

Las edades, tienen una distribución asimétrica a la derecha.

## 6.3. COEFICIENTE DE CURTOSIS

La curtosis es la deformación vertical (apuntamiento) de una distribución de frecuencias. Este coeficiente determina el grado de concentración de los datos con respecto a la media aritmética; se calcula cuando la distribución de frecuencias tiene la forma simétrica (forma de campana). El coeficiente de curtosis se calcula por la siguiente fórmula:

Para datos no agrupados

$$k = \frac{\sum_{i=1} (x_i - \bar{x})^4}{ns^4} - 3$$

Para datos discretos agrupados

$$k = \frac{\sum_{i=1} (x_i - \bar{x})^4 f_i}{ns^4} - 3$$

Para datos continuos agrupados

$$k = \frac{\sum_{i=1}^{k} (x'_i - \bar{x})^4 f_i}{ns^4} - 3$$

$k = 0$ , la distribución será mesocúrtica o normal, concentración normal de los datos con respecto a la media aritmética.

$k > 0$ , la distribución será leptocúrtica o mas apuntada que la normal, alta concentración de los datos con respecto a la media aritmética

$k < 0$, la distribución será platicúrtica o menos apuntada que la normal. baja concentración de los datos con respecto a la media aritmética

Otro coeficiente de curtosis en función de cuartiles y percentiles es:

$$k = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

$k = 0.263$, la distribución de frecuencias será mesocúrtica o normal

$k > 0.263$, la distribución de frecuencias será Leptocúrtica más apuntada que la normal, los datos se encuentran más concentrados a la media aritmética $k < 0.263$, la distribución de frecuencias será Platicúrtica, menor apuntada que la normal, los datos se encuentran menos concentrados a la media aritmética.

Ejemplo aplicativo:

La tabla de distribución de frecuencias, corresponde a las edades (años) de 42 usuarios que concurren a un centro de información.

| Edades(años) | $f_i$ | $x_i'$ |
|---|---|---|
| 18.5-22.5 | 3 | 20.5 |
| 22.5-26.5 | 4 | 24.5 |
| 26.5-30.5 | 14 | 28.5 |
| 30.5-34.5 | 12 | 32.5 |
| 34.5-38.5 | 5 | 36.5 |
| 38.5-42.5 | 4 | 40.5 |
| Total | 42 | |

¿A que tipo de apuntamiento corresponde la distribución de las edades de los usuarios?.

Cálculos:

$$\sum_{i=1}^{6}(x_i' - \bar{x})^4 f_i = 81,260.2216$$

$$\sum_{i=1}^{6}(x_i' - \bar{x})^2 f_i = 1124.57143$$

$$k = \frac{42(81260.2216)}{(1124.57143)^2} - 3 = 2.698691 - 3 = -0.301309 < 0$$

La distribución de frecuencias de las edades será platicurtica.

Usando el otro coeficiente se tiene:

De la tabla de distribución frecuencias se tiene:

$Q_3 = 34$

$Q_1 = 27.5$

$P_{90} = 38.34$

$P_{10} = 23.7$

$$k = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{34 - 27.5}{2(38.34 - 23.7)} = 0.22199$$

La distribución de las edades será platicurtica; esto implica que las edades de los usuarios están poco concentrados con respecto a la media aritmética.

Otra fórmula para calcular el coeficiente de curtosis:

$$k = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Siendo: $\bar{x}$ y $s$ la media aritmética y la desviación estándar respectivamente.

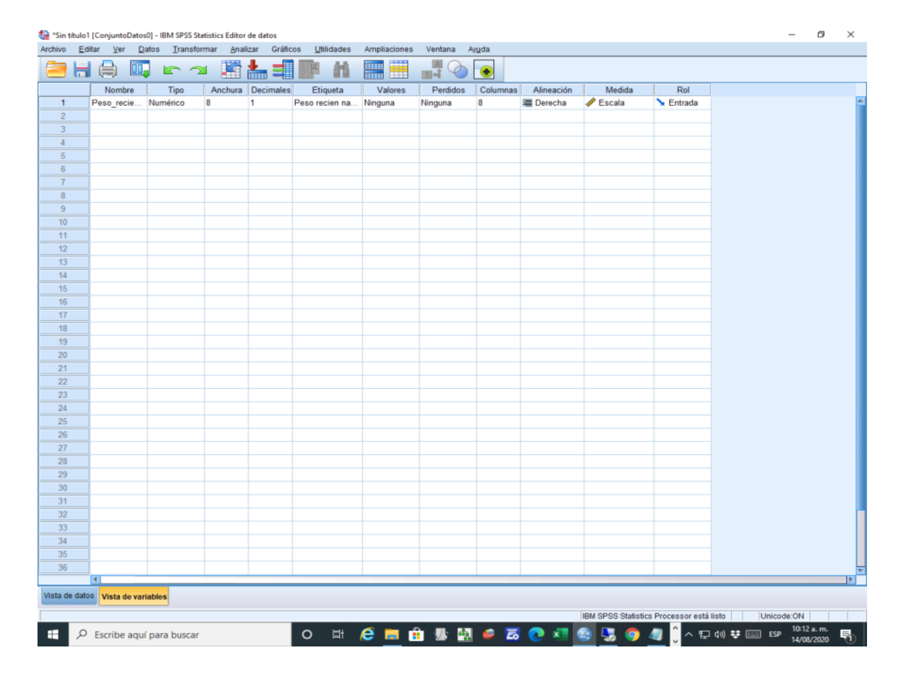
$k = 0$ , la distribución será mesocúrtica o normal, media concentración de los datos con respecto a la media aritmética.

$k > 0$ , la distribución será leptocúrtica o mas apuntada que la normal, alta concentración de los datos con respecto a la media aritmética

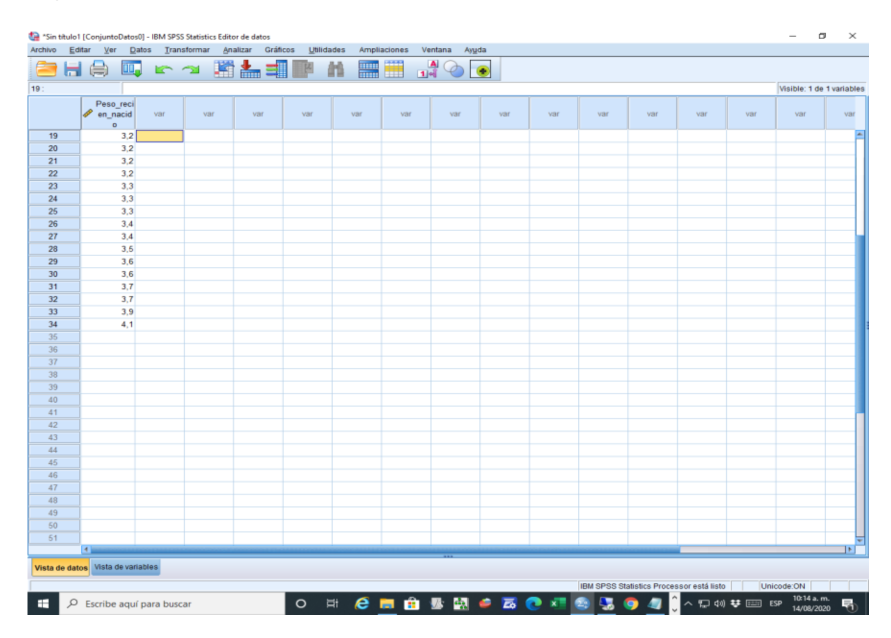
$k < 0$, la distribución será platicúrtica o menos apuntada que la normal. baja concentración de los datos con respecto a la media aritmética

Aplicando el SPSS para realizar el cálculo de la simetría, asimetría, curtosis

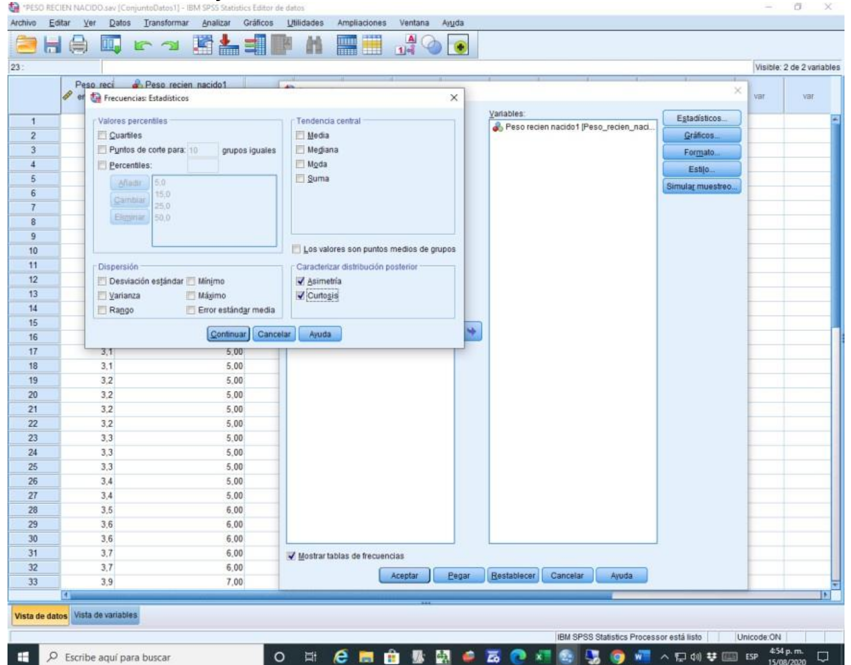
Definiendo las variables




## Ingresando los datos



En el Menú del SPSS escoger ANALIZAR-ESTADÍSTICOS DESCRIPTIVOS—FRECUENCIA-pasar la variable Peso_recien_nacido a la casilla de la derecha-ESTADISTICOS-caracterización posterior marcar asimetría y curtosis-CONTINUAR-ACEPTAR




**Estadísticos**

Peso recien nacido1

| | | |
|---|---|---|
| N | Válido | 34 |
| | Perdidos | 0 |
| Asimetría | | -,308 |
| Error estándar de asimetría | | ,403 |
| Curtosis | | -,728 |
| Error estándar de curtosis | | ,788 |

## 6.4. Ejercicios propuestos

1. En una agencia bancaria, el tiempo (en minutos) de atención a los 70 clientes se presenta en la siguiente tabla de distribución de frecuencia:

| Tiempo (minutos) | $f_i$ |
|---|---|
| [1.5 2.0) | 2 |
| [2.0 2.5) | 6 |
| [2.5 3.0) | 12 |
| [3.0 3.5) | 30 |
| [3.5 4.0) | 12 |
| [4.0 4.5) | 6 |
| [4.5 5.0) | 2 |
| Total | 70 |

Hallar e interprete el coeficiente de asimetría y el coeficiente de curtosis.

2. Los datos que a continuación se presentan corresponden a edades de 60 usuarios que concurren a un centro de información (biblioteca)

| Edad(años) | $f_i$ |
|---|---|
| 18.5-22.5 | 4 |
| 22.5-26.5 | 24 |
| 26.5-30.5 | 15 |
| 30.5-34.5 | 10 |
| 34.5-38.5 | 5 |
| 38.5-42.5 | 2 |
| Total | 60 |

Hallar e interprete el coeficiente de asimetría.

3. Los datos que a continuación se presentan corresponden a edades de usuarios que concurren a un centro de información (biblioteca)

| Edad(años) | $h_i$ |
|---|---|
| 12-14 | 0.02 |
| 14-16 | 0.2 |
| 16-18 | 0.56 |
| 18-20 | 0.2 |
| 20-22 | 0.02 |
| Total | 1.0 |

Hallar e interprete el coeficiente de curtosis.

4. Los datos que a continuación se presentan corresponden a las estaturas (mts.) de 44 alumnos elegidos al azar, de la universidad Nacional Mayor de San Marcos.

| Estaturas (mts) | $f_i$ |
|---|---|
| 1.65-1.68 | 3 |
| 1.68-1.71 | 10 |
| 1.71-1.74 | 17 |
| 1.74-1.77 | 8 |
| 1.77-1.80 | 4 |
| 1.80-1.83 | 2 |
| Total | 44 |

Hallar e interprete el coeficiente de asimetría.

5. Los datos que a continuación se presentan corresponden a montos por ventas (miles de nuevos soles) de 60 establecimientos comerciales de Lima Metropolitana.

| Montos de ventas | $f_i$ |
|---|---|
| 12.5-17.5 | 2 |
| 17.5-22.5 | 5 |
| 22.5-27.5 | 23 |
| 27.5-32.5 | 23 |
| 32.5-37.5 | 5 |
| 37.5-42.5 | 2 |
| Total | 60 |

Hallar e interprete el coeficiente de curtosis.

# CAPÍTULO 7
# VARIABLES BIDIMENSIONALES

## INTRODUCCIÓN

En esta CAPÍTULO se estudiaran las variables bidimensionales, las cuales se presentan cuando se observan en forma simultánea dos variables $(X;Y)$ en la unidad elemental, en una población o muestra aleatoria.

## 7.1. VARIABLES BIDIMENSIONALES

Las variables bidimensionales $(X, Y)$ pueden ser ambas cualitativas o cuantitativas o bien una mezcla de ambas. Cualquiera que sea el caso los datos se clasifican bajo dos criterios y se presentan en tabla de doble entrada o tabla de contingencia.

Ejemplos:

- Estaturas y pesos de los alumnos de la Universidad Nacional Mayor de San Marcos
- Sexo y Opinión de los encuestados
- Número de heridos y causas de los accidentes de transito
- Gasto e ingreso de los jefes de familia
- Altura y diámetro de los árboles
- Grado de instrucción y adaptabilidad al matrimonio de los cónyuges
- Número de personas y cantidad de habitaciones en una casa

Supongamos que la variable X se clasifica en “m” clases y la variable Y en “k” clases. Cada $f_{ij}$ nos indica el número de veces que la unidad elemental presenta la clase $X_i$ e $Y_j$ simultáneamente, obteniéndose de esta forma una distribución conjunta de las variables X e Y. Ver tabla 01

Tabla 01

| $Y$ \ $X$ | $Y_1$ | $Y_2$ | . . . | $Y_k$ | Total |
|---|---|---|---|---|---|
| $X_1$ | $f_{11}$ | $f_{12}$ | . . . | $f_{1k}$ | $f_{1.}$ |
| $X_2$ | $f_{21}$ | $f_{22}$ | . . . | $f_{2k}$ | $f_{2.}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $X_m$ | $f_{m1}$ | $f_{m2}$ | . . . | $f_{mk}$ | $f_{m.}$ |
| Total | $f_{.1}$ | $f_{.2}$ | . . . | $f_{.k}$ | $f_{..}$ |

Ejemplo

De 98 Tesis de Física elegidas al azar, se observó simultáneamente dos variables de interés:
X = Sexo del autor de la tesis
Y = Calificativo en la sustentación

Tabla 02

Distribución conjunta de las frecuencias de las variables, sexo del autor de la tesis (X) y calificativo en la sustentación (Y)

| Calificativo \ Sexo | Excelente | Muy Bueno | Bueno | Regular | Total |
|---|---|---|---|---|---|
| Varón | 8 | 12 | 15 | 12 | 47 |
| Dama | 6 | 10 | 20 | 15 | 51 |
| Total | 14 | 22 | 35 | 27 | 98 |

Fuente: Biblioteca Central de la Universidad Nacional Mayor de San Marcos

Interpretación:

$f_{11}$ =8 (Significa que existen 8 tesis de física que corresponden a varones y obtuvieron el calificativo de excelente).

$f_{1.}$ =47 (Significa que existen 47 tesis de física, cuyos autores son varones,).

$f_{.2}$ =22(Significa que existen 22 tesis de física que obtuvieron el calificativo de muy bueno.

La Tabla 02, también puede presentarse por sus frecuencias relativas para cada casillero.

Tabla 03

Distribución Conjunta de las frecuencias relativas porcentuales de las variables, sexo del autor de la tesis (X), y calificativo en la sustentación (Y)

| Calificativo Sexo | Excelente | Muy Bueno | Bueno | Regular | Total |
|---|---|---|---|---|---|
| Varón | 8.16% | 12.24% | 15.31% | 12.24% | 47.96% |
| Dama | 6.12% | 10.20% | 20.41% | 15.31% | 52.04% |
| Total | 14.29% | 22.45% | 35.71% | 27.55% | 100.00 % |

Fuente: Biblioteca Central de la Universidad Nacional Mayor de San Marcos

## 7.2. DISTRIBUCIONES MARGINALES

Si de la tabla 01, consideramos sólo los totales para las categorías de $X$ , sin tener en cuenta las categorías de Y, se obtiene la distribución marginal de $X$ ; es decir:

TABLA DE DISTRIBUCIÓN MARGINAL DE $X$

| $X$ | $f_{i.}$ |
|---|---|
| $X_1$ | $f_{1.}$ |
| $X_2$ | $f_{2.}$ |
| . | . . |
| $X_m$ | $f_{m.}$ |
| **Total** | $f_{..}$ |

De la misma forma la distribución marginal para $Y$ será:

TABLA DE DISTRIBUCIÓN MARGINAL DE $Y$

| $Y$ | $f_{.j}$ |
|---|---|
| $Y_1$ | $f_{.1}$ |
| $Y_2$ | $f_{.2}$ |
| | |
| $Y_k$ | $f_{.k}$ |
| **Total** | $f_{..}$ |

## 7.3. DISTRIBUCIONES CONDICIONALES

Se obtiene al poner una restricción o condición a una de las dos variables; es decir:

Distribución condicional de $X$ dado $Y$ , se simboliza por: $X/Y = Y_j$

TABLA DE DISTRIBUCIÓN CONDICIONAL DE $X/Y = Y_j$

| $X/Y = Y_j$ | $Y_j$ |
|---|---|
| $X_1$ | $f_{1j}$ |
| $X_2$ | $f_{2j}$ |
| . | . |
| $X_m$ | $f_{mj}$ |
| **Total** | $f_{.j}$ |

También la distribución condicional de $Y$ dado $X$ , se simboliza por: $Y/X = X_i$

TABLA DE DISTRIBUCIÓN CONDICIONAL DE $Y/X = X_i$

| $Y/X = X_i$ | $X_i$ |
|---|---|
| $Y_1$ | $f_{i1}$ |
| $Y_2$ | $f_{i2}$ |
| . . | |
| $Y_k$ | $f_{ik}$ |
| Total | $f_{i.}$ |

## 7.4. INDEPENDENCIA ESTADÍSTICA

Se dice que dos variables $X$ e $Y$ , son independientes estadísticamente, cuando las frecuencias relativas conjuntas es igual al producto de las frecuencias relativas marginales para todas las celdas de la tabla 01; es decir,

$$\frac{f_{ij}}{n} = \frac{f_{i.}}{n} * \frac{f_{.j}}{n} \quad \forall_{ij}$$

Si esta condición no se cumple para todas las celdas, se dice que hay dependencia estadística.

Medias y Variancias Marginales:

De la tabla 01 sea la variable bidimensional $(X,Y)$ . Las medias y variancias marginales de $X$ e $Y$ , son dadas por:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_{i.}}{n}$$

$$s_x^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_{i.}}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^{k} y_j f_{.j}}{n}$$

$$s_y^2 = \frac{\sum_{i=1}^{r} (y_j - \bar{y})^2 f_{.j}}{n}$$

Siendo: $n = \sum_{i=1}^{k}\sum_{j=1}^{r} f_{ij} \begin{cases} i = 1,2,3,...k \\ j = 1,2,3,...r \end{cases}$

## 7.5. COVARIANCIA

Es el estudio de la variabilidad conjunta de dos variables aleatoria, la medida que se utiliza para cuantificar esta relación es:

$$S_{xy} = Cov(x, y) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{r}(x_i - \bar{x})(y_j - \bar{y}) f_{ij}}{n} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{r} f_{ij} x_i y_j - n\overline{xy}}{n}$$

$$S_{xy} = \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{r} f_{ij} x_i y_j\right) - \overline{xy}$$

$S_{xy} > 0 \rightarrow$ Hay dependencia lineal directa (positiva); es decir a grandes valores de $X$, corresponden grandes valores de $Y$.

$S_{xy} = 0 \rightarrow$ No hay dependencia lineal entre las variables (están incorrelacionadas); es decir no hay relación lineal

$S_{xy} < 0 \rightarrow$ Hay dependencia lineal inversa o negativa; es decir a grandes valores de $X$, corresponden pequeños valores de $Y$.

- Propiedades de la Covariancia

1. Si a todos los valores de la variable $X$ e $Y$, les sumamos una constante $k$; la covariancia original no cambia.
2. Si a todos los valores de la variable $X$ e $Y$, les multiplicamos por una constante $k_1$ y $k_2$, respectivamente, la

covariancia original queda multiplicada por el producto de las constantes.

3. Sea $S_{xy}$, la covariancia de la variable bidimensional $X\ e\ Y$, sean las transformaciones lineales para cada variable $z = ax + b\ ,\ w = cy + d$, la nueva covarianza es dada por: $S_{zw} = acS_{xy}$

Coeficiente de correlación lineal de X e Y

El inconveniente de la covariancia, como medida de asociación es su dependencia de las unidades de medidas que utiliza. Para salvar esta dificultad, se define otra medida de la asociación como la correlación, que no está afectada por las unidades de medidas.

$$r_{xy} = \frac{Cov(x, y)}{S_x.S_y}$$

Ejemplo aplicativo

Sea la tabla de datos

Tabla 01

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| N° de hijos(x) | 2 | 3 | 4 | 2 | 4 | 3 | 5 | 2 | 2 | 3 |
| N° de dormitorios (Y) | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |

Elaborar la tabla de distribución de frecuencias de X e Y

Tabla 02

| (Y) / (X) | 1 | 2 | Total |
|---|---|---|---|
| 2 | 3 | 1 | 4 |
| 3 | 2 | 1 | 3 |
| 4 | 0 | 2 | 2 |
| 5 | 0 | 1 | 1 |
| Total | 5 | 5 | 10 |

Hallar las distribuciones marginales de X e Y

Distribución marginal de X

Tabla 03

| $X$ | $f_x$ |
|---|---|
| 2 | 4 |
| 3 | 3 |
| 4 | 2 |
| 5 | 1 |
| Total | 10 |

Distribución marginal de Y

Tabla 04

| $Y$ | $f_y$ |
|---|---|
| 1 | 5 |
| 2 | 5 |
| Total | 10 |

Distribución condicional de Y dado X=3

Tabla 05

| $Y$ | $f(y/x=3)$ |
|---|---|
| 1 | 2 |
| 2 | 1 |
| Total | 3 |

Calcular la covariancia

Utilizando la formula de la covariancia, con los datos de la tabla 02

$$S_{xy} = \frac{1}{10}\left(\sum_{i=1}^{4}\sum_{j=1}^{2} f_{ij}x_i y_j\right) - \bar{x}\bar{y}$$

$$S_{xy} = \frac{1}{10}((3)(2)(1) + (1)(2)(2) + (2)(3)(1) + (1)(3)(2) + (2)(4)(2) + (1)(5)(2) - 3(1.5)$$

$$S_{xy} = 0.3$$

Calcular la variancia marginal de X

De la tabla 03

$$\sum_{i=1}^{3} x_i f_i = 2(4) + 3(3) + 4(2) + 5(1) = 30$$

$$\sum_{i=1}^{3} x_i^2 f_i = 2^2(4) + 3^2(3) + 4^2(2) + 5^2(1) = 100$$

$$S_x^2 = \frac{100 - \frac{(30)^2}{10}}{10} = 1$$

Calcular la variancia marginal de Y

De la tabla 04

$$\sum_{i=1}^{2} y_i f_i = 1(5) + 2(5) = 15$$

$$\sum_{i=1}^{2} y_i^2 f_i = 1^2(5) + 2^2(5) = 25$$

$$S_y^2 = \frac{25 - \frac{(15)^2}{10}}{10} = 0.25$$

Calcular el coeficiente de correlación de X e Y

Utilizando la formula de la correlación y los cálculos obtenido anteriormente

$$r_{xy} = \frac{S_{xy}}{S_x . S_y} = \frac{0.3}{\sqrt{1}.\sqrt{0.25}} = 0.6$$

De otra forma de la tabla 01

$$\sum_{i=1}^{10} x_i y_i = 48$$

$$\sum_{i=1}^{10} x_i = 30$$

$$\sum_{i=1}^{10} x_i^2 = 100$$

$$\sum_{i=1}^{10} y_i = 15$$

$$\sum_{i=1}^{10} y_i^2 = 25$$

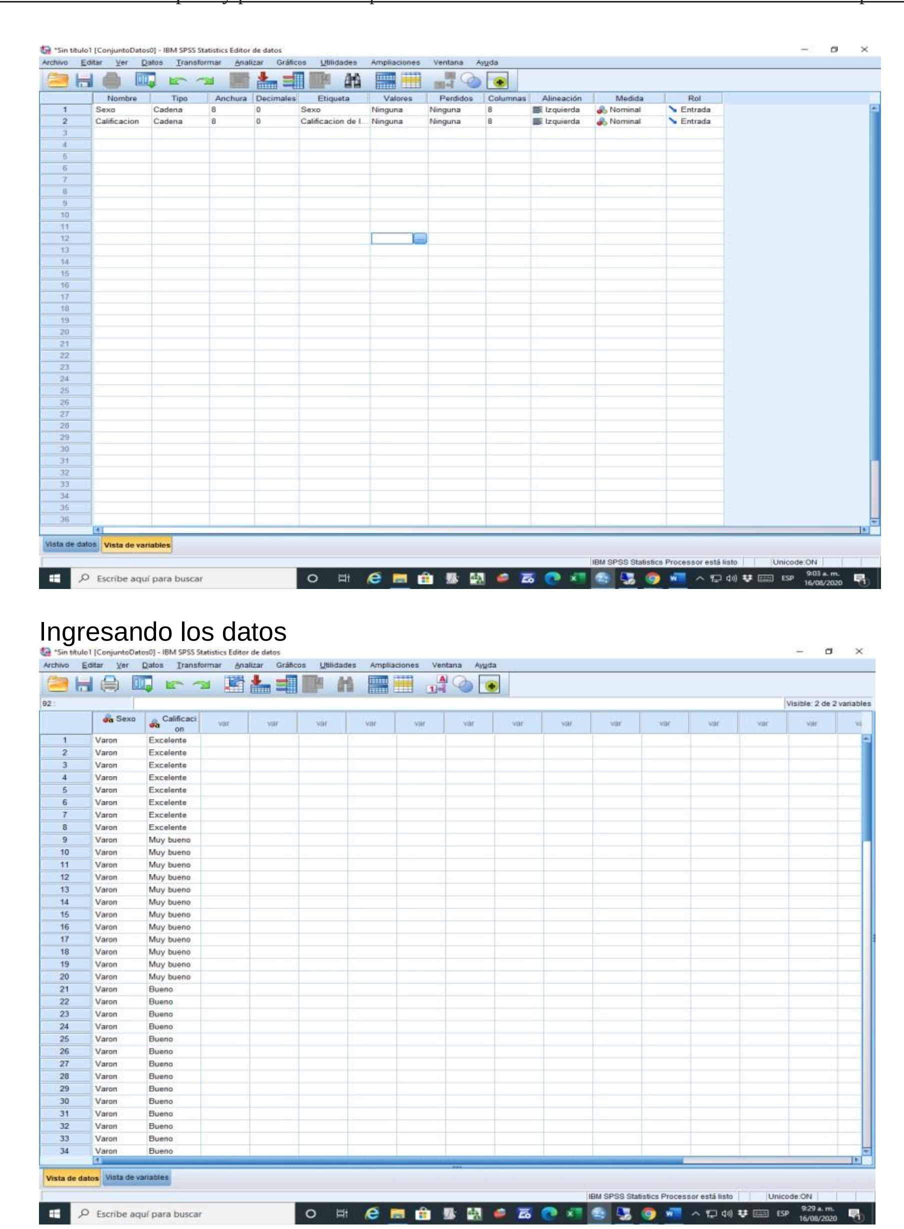
El coeficiente de correlación de Xe Y

$$r_{xy} = \frac{\sum_{i=1}^{10} x_i y_i - \frac{(\sum_{i=1}^{10} x_i)(\sum_{i=1}^{10} y_i)}{10}}{\sqrt{\sum_{i=1}^{10} x_i^2 - \frac{(\sum_{i=1}^{10} x_i)^2}{10}}\sqrt{\sum_{i=1}^{10} y_i^2 - \frac{(\sum_{i=1}^{10} y_i)^2}{10}}} = \frac{48 - \frac{(30)(15)}{10}}{\sqrt{100 - \frac{(30)^2}{10}}\sqrt{25 - \frac{(15)^2}{10}}}$$

$$r_{xy} = \frac{3}{\sqrt{10}.\sqrt{2.5}} = \frac{3}{\sqrt{25}} = \frac{3}{5} = 0.6$$

Aplicando el SPSS para realizar tablas cruzadas o de contingencia

Definiendo las variables




En el Menú del SPSS escoger ANALIZAR-ESTADISTICOS DESCRIPTIVOS—TABLAS CRUZADAS-pasar la variable sexo a la

casilla filas y la variable Calificacion de la tesis la columnas -CASILLAS-CONTINUAR-ACEPTAR

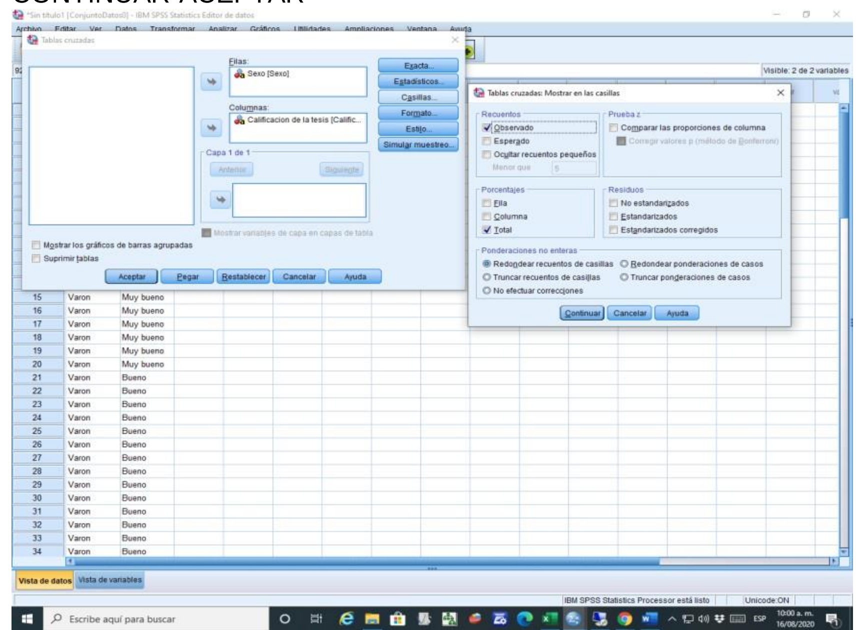



**Tabla cruzada Sexo*Calificacion de la tesis**

| | | | Calificacion de la tesis | | | | Total |
|---|---|---|---|---|---|---|---|
| | | | Bueno | Excelente | Muy bueno | Regular | |
| Sexo | Dama | Recuento | 20 | 6 | 10 | 15 | 51 |
| | | % del total | 20,4% | 6,1% | 10,2% | 15,3% | 52,0% |
| | Varon | Recuento | 15 | 8 | 12 | 12 | 47 |
| | | % del total | 15,3% | 8,2% | 12,2% | 12,2% | 48,0% |
| Total | | Recuento | 35 | 14 | 22 | 27 | 98 |
| | | % del total | 35,7% | 14,3% | 22,4% | 27,6% | 100,0% |

## 7.6. Ejercicio propuesto:

a. Elaborar una tabla de doble entrada con los datos que se presentan a continuación.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | V | D | V | D | V | D | D | D | V | D | V | D | D | D |
| Candidato | A | C | B | A | A | B | A | B | A | A | C | A | B | A |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | D | V | V | D | V | V | D | D | V | D | D | V | D | D |
| Candidato | B | A | B | A | A | B | A | C | A | A | B | A | A | A |

Con la información elaborar la tabla de doble entrada e interpretar algunos resultados.

b. Los siguientes datos corresponden a una encuesta tomada a 56 alumnos de la Universidad Nacional Mayor de San Marcos, con la finalidad de evaluar los servicios que prestan a los usuarios la biblioteca Central.

c.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | V | D | V | V | V | D | D | D | V | D | V | D | D | D |
| Servicio Del CI | R | R | B | R | M | R | B | M | R | R | B | B | B | R |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | V | D | V | D | V | D | D | D | V | D | V | D | D | D |
| Servicio Del CI | B | M | R | R | B | B | M | B | M | R | R | B | B | B |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | V | D | V | D | V | D | D | D | V | D | V | D | D | D |
| Servicio Del CI | R | B | R | R | B | R | R | B | R | R | B | R | M | R |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | V | D | V | D | V | D | D | D | V | D | V | D | D | D |
| Servicio Del CI | B | B | B | M | R | B | M | B | B | R | M | R | B | M |

LEYENDA
SEXO: V = VARON M= MUJER
SERVICIOS DEL CI (Centro de Información): M=MALO
R=REGULAR B= BUENO
Con la información elaborar la tabla de doble entrada e interpretar algunos resultados.

# CAPÍTULO 8

## INTRODUCCIÓN

En este capítulo, se estudia la relación lineal o no lineal entre dos variables $(X, Y)$ con el objetivo de conocer si estas están asociadas o dependientes

## 8.1. COEFICIENTRE DE CORRELACIÓN DE PEARSON

Es un valor que mide el grado de asociación lineal entre dos variables cuantitativas $(X, Y)$

Sean $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$, . . . , $(x_n, y_n)$ datos de dos variables de una muestra aleatoria de tamaño $n$ extraídas de una población. El coeficiente se calcula por la fórmula:

$$r_p = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \cdot \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}}}$$

Si $r_p = 0$ , entonces los valores que toma la variable $(X, Y)$ , no están asociadas linealmente.

Si $r_p \to 1$ , entonces los valores que toma la variable $(X, Y)$ , están asociadas linealmente; es decir, si $X$ crece, también $Y$ crece (viceversa).

Si $r_p \to -1$, entonces los valores que toma la variable $(X, Y)$ , están asociadas linealmente; es decir, si $X$ crece $Y$ decrece (viceversa).

Ejemplo
Se eligen al azar 10 estudiantes y se registran sus notas de estadística I y estadística II, tal como se presenta en el cuadro:

| $X$ | $Y$ |
|---|---|
| 13 | 13 |
| 12 | 12 |
| 14 | 15 |
| 15 | 15 |
| 16 | 17 |
| 14 | 16 |
| 13 | 14 |
| 15 | 16 |
| 11 | 12 |
| 13 | 15 |

¿Se encuentran los datos correlacionados?

Calculando las sumatorias

$\sum_{i=1}^{10} x_i y_i = 1993$ $\quad \sum_{i=1}^{10} x_i = 136$ $\quad \sum_{i=1}^{10} y_i = 145$ $\quad \sum_{i=1}^{10} x_i^2 = 1870$ $\quad \sum_{i=1}^{10} y_i^2 = 2129$

$$r_p = \frac{1993 - \frac{(136)(145)}{10}}{\sqrt{1870 - \frac{136^2}{10}} \cdot \sqrt{2129 - \frac{145^2}{10}}} = 0.9032$$

Conclusión: las notas de los estudiantes obtenidas en el curso de estadística I y II están asociadas o correlacionadas linealmente

## 8.2. COEFICIENTE DE CORRELACIÓN DE RANGOS DE SPEARMAN

Es un valor que mide el grado de asociación lineal o no lineal entre dos variable cualitativas ordinales o jerárquicas $(X, Y)$, los datos de ambas variables se ordenan por rangos creciente o decreciente

Para una muestra aleatoria de tamaño $n$, sea $(x_1, x_2, x_3, \ldots x_n)$ los datos de la primera variable con sus rangos correspondientes $(x_1^r, x_2^r, x_3^r, \ldots, x_n^r)$; para los valores de la segunda variable

$(y_1, y_2, y_3,... y_n)$ con sus rangos correspondientes $(y_1^r, y_2^r, y_3^r,..., y_n^r)$ .

El coeficiente se calcula por la fórmula:

$$r_s = 1 - \frac{6.\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

Donde: $d_i = x_i^r - y_i^r$

Si $r_p = 0$ , entonces los valores que toma la variable $(X, Y)$ , no están asociadas.

Si $r_p \to 1$ , entonces los valores que toma la variable $(X, Y)$ , están asociadas; es decir, si $X$ crece también $Y$ también crece (viceversa).

Si $r_p \to -1$, entonces los valores que toma la variable $(X, Y)$ , están asociadas; es decir, si $X$ crece $Y$ decrece (viceversa).

Nota

Cuando hay empates de rangos, tanto para los valores de $X$ como $Y$ , al momento de asignarle sus rangos, se asigna a las observaciones empatadas la media aritmética de los rangos se cumple que

Ejemplo

| $X$ | 7 | 6 | 5 | 6 | 4 |
|---|---|---|---|---|---|
| | 5 | 3 | 2 | 4 | 1 |
| $r$ | 5 | 3.5 | 2 | 3.5 | 1 |

Nota

Cuando no hay empates de rangos, tanto para los valores de $X$ como $Y$ , se cumple que el coeficiente de correlación de Spearman es igual al coeficiente de correlación de Pearson calculado por los rangos.

$$r_s = r_p$$

Nota

Cuando hay algunos empates de rangos, tanto para los valores de $X$ como $Y$ , el coeficiente de correlación de Spearman es

aproximadamente igual al coeficiente de correlación de Pearson calculado por los rangos.

$$r_s \approx r_p$$

Ejemplo

Se eligen al azar 7 estudiantes y se registran las $X$ =horas de estudios dedicados a la asignatura de matemática con sus correspondientes $Y$ = notas obtenidas a la referida asignatura:

| $X$ | $Y$ |
|---|---|
| 8 | 12 |
| 5 | 11 |
| 11 | 14 |
| 13 | 15 |
| 10 | 13 |
| 5 | 12 |
| 8 | 13 |

¿Calcular el índice de correlación de Spearman?

Se calcula los rangos para cada variable

| $X$ | $Y$ | $x_i^r$ | $y_i^r$ | $d_i = x_i^r - y_i^r$ | $d_i^2$ |
|---|---|---|---|---|---|
| 8 | 12 | 3.5 | 2.5 | 1 | 1 |
| 5 | 11 | 1.5 | 1 | 0.5 | 0.25 |
| 11 | 14 | 6 | 6 | 0 | 0 |
| 13 | 15 | 7 | 7 | 0 | 0 |
| 10 | 13 | 5 | 4.5 | 0.5 | 0.25 |
| 5 | 12 | 1.5 | 2.5 | -1 | 1 |
| 8 | 13 | 3.5 | 4.5 | -1 | 1 |
| | | | | total | 3.5 |

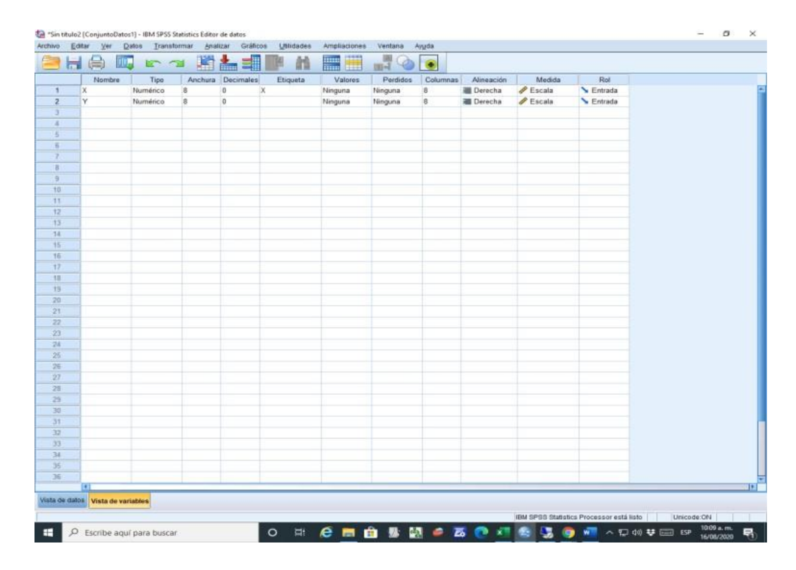
Usando la formula

$$r_s = 1 - \frac{6.\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6.\sum_{i=1}^{7} d_i^2}{7(7^2-1)} = 1 - \frac{6(3.5)}{7(49-1)} = 0.9375$$

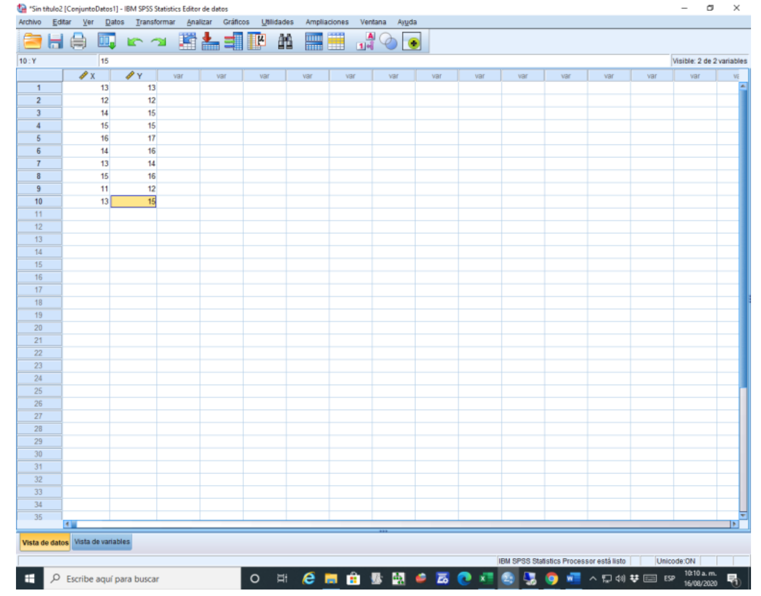
Conclusión: las horas de estudio dedicadas a la asignatura matemática, están asociadas a las notas obtenidas.

## Aplicando el SPSS para realizar el Índice de correlación de Pearson

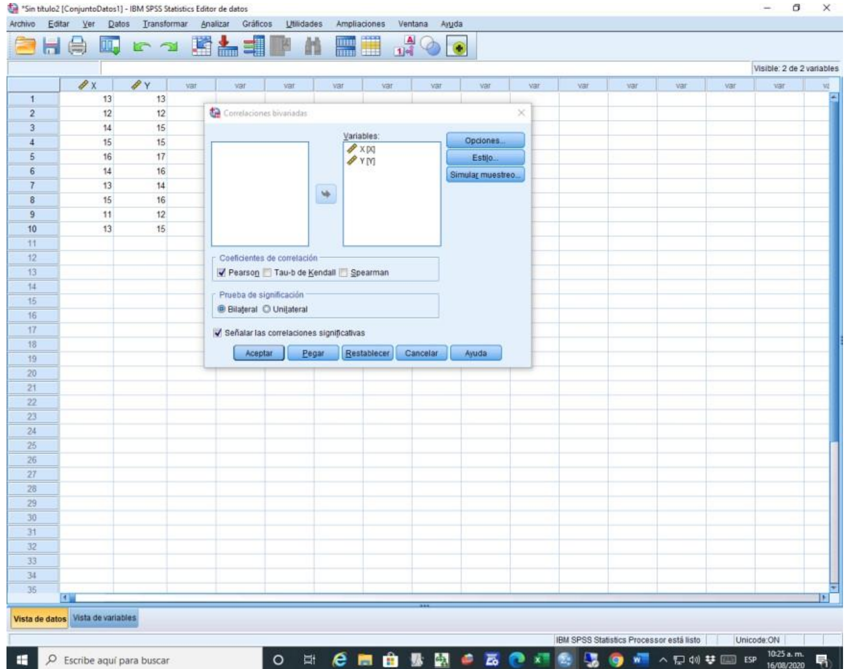
### Definiendo las variables




### Ingresando los datos



En el Menú del SPSS escoger ANALIZAR-ESTADÍSTICOS DESCRIPTIVOS—TABLAS CRUZADAS-pasar la variable X y la variable Y al casillero de la derecha, marcar Pearson-ACEPTAR

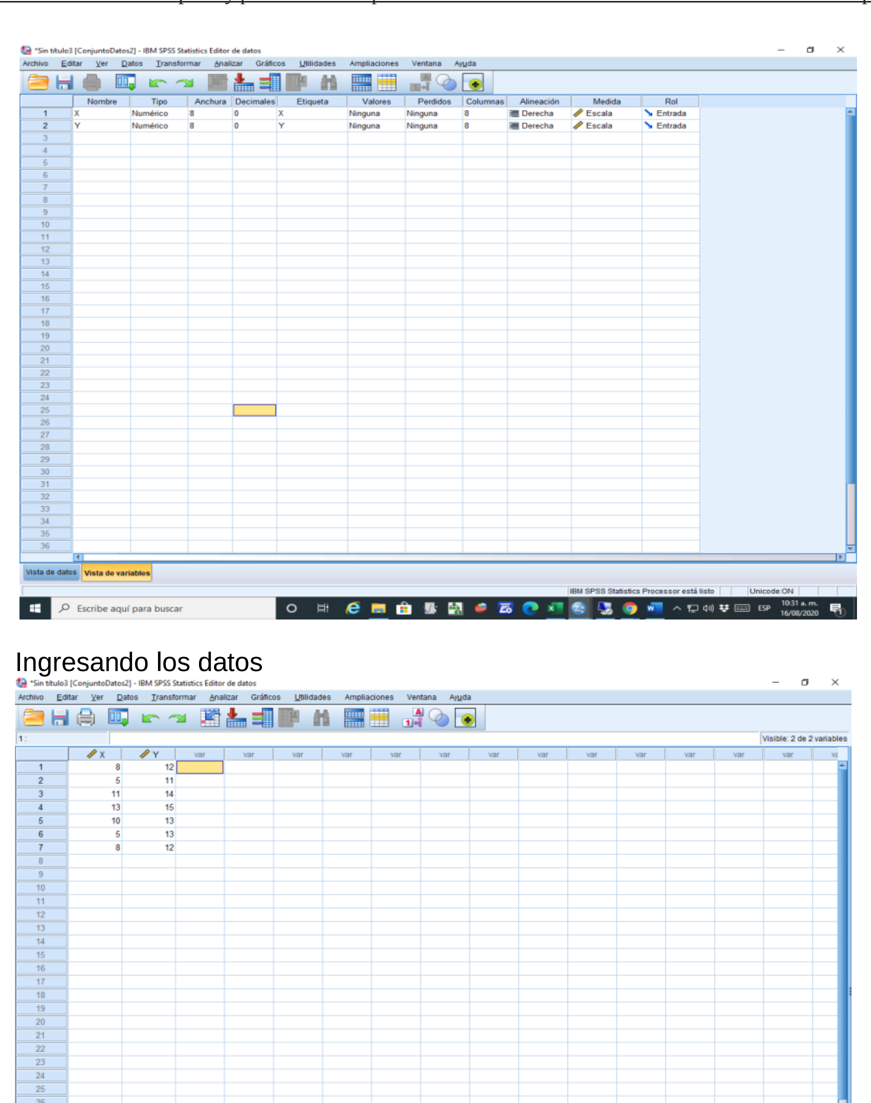


**Correlaciones**

| | | X | Y |
|---|---|---|---|
| X | Correlación de Pearson | 1 | ,903** |
| | Sig. (bilateral) | | ,000 |
| | N | 10 | 10 |
| Y | Correlación de Pearson | ,903** | 1 |
| | Sig. (bilateral) | ,000 | |
| | N | 10 | 10 |

**. La correlación es significativa en el nivel 0,01 (bilateral).

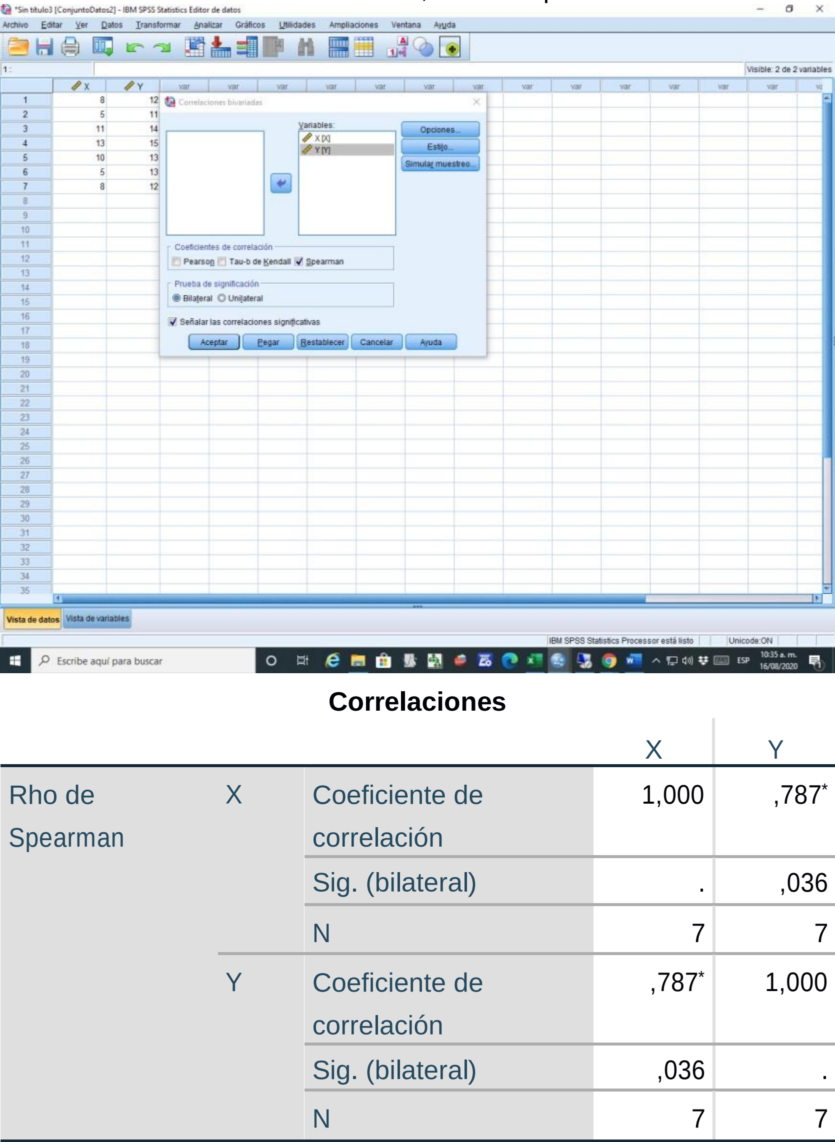
Aplicando el SPSS para realizar el Indice de correlación de Spearman Definiendo las variables




## Ingresando los datos

| | X | Y |
|---|---|---|
| 1 | 8 | 12 |
| 2 | 5 | 11 |
| 3 | 11 | 14 |
| 4 | 13 | 15 |
| 5 | 10 | 13 |
| 6 | 5 | 13 |
| 7 | 8 | 12 |

En el Menú del SPSS escoger ANALIZAR-ESTADÍSTICOS DESCRIPTIVOS—TABLAS CRUZADAS-pasar la variable X y la variable Y al casillero de la derecha, marcar Spearman-ACEPTAR




**Correlaciones**

| | | | X | Y |
|---|---|---|---|---|
| Rho de Spearman | X | Coeficiente de correlación | 1,000 | ,787* |
| | | Sig. (bilateral) | . | ,036 |
| | | N | 7 | 7 |
| | Y | Coeficiente de correlación | ,787* | 1,000 |
| | | Sig. (bilateral) | ,036 | . |
| | | N | 7 | 7 |

*. La correlación es significativa en el nivel 0,05 (bilateral).

# CAPÍTULO 9
# TÉCNICAS DEL CONTEO

## INTRODUCCIÓN

El objetivo de este capítulo es presentar las formulas y técnicas que permiten determinar el total de grupos que se pueden formar con los elementos de uno o más conjuntos. Los grupos formados pueden ser sin reemplazo o con reemplazo y a la vez tomando en cuenta o no el orden de sus elementos.

Sean los dígitos $\{1, 2, 3\}$, se eligen 2 dígitos para formar un número, determinar los números formados:

12 13 23 — sin reemplazo y sin orden

12 13 23
21 31 32 — sin reemplazo y con orden

12 13 23
11 22 33 — con reemplazo y sin orden

12 13 23
21 31 32
11 22 33 — con reemplazo y con orden

Orden, es permutar los elementos dentro del grupo
Sin reemplazo, el elemento elegido aparece una solo vez en el grupo
Con reemplazo, el elemento elegido aparece más de una vez en el grupo

9.1. Con un solo conjunto

- Si se tiene un conjunto con "n" elementos diferentes, todos los posibles grupos que se pueden formar, si se

toman “r” elementos (para $r \leq n$), sin tomar en cuenta el orden y sin reemplazo, es dado por:

$$C_r^n = \frac{n!}{r!(n-r)!}$$

- Si se tiene un conjunto con “n” elementos diferentes, todos los posibles grupos que se pueden formar, si se toman “r” elementos (para $r \leq n$), sin tomar en cuenta el orden y con reemplazo, es dado por:

$$C_r^{n+r-1} = \frac{(n+r-1)!}{r!(n-1)!}$$

- Si se tiene un conjunto con “n” elementos diferentes, todos los posibles grupos que se pueden formar, si se toman “r” elementos (para $r \leq n$), tomando en cuenta el orden y sin reemplazo, es dado por:

$$V_r^n = \frac{n!}{(n-r)!}$$

Un caso particular, es cuando $r = n$, luego $V_n^n = n!$

- Si se tiene un conjunto con “n” elementos diferentes, todos los posibles grupos que se pueden formar, si se toman “r” elementos (para $r \leq n$), tomando en cuenta el orden y con reemplazo, es dado por:

$$n^r$$

- Si se tiene un conjunto con “n” elementos de los cuales $n_1$ son iguales, $n_2$ son iguales,..., $n_p$ son iguales, donde se cumple que $\sum_{i=1}^{p} n_i = n$, luego el total de permutaciones para los “n” elementos es:

$$\frac{n!}{(n_1!)(n_2!)...(n_p!)}$$

## 9.2. Con dos o más conjuntos

- Principio de la multiplicación

Un acontecimiento puede ocurrir de “a “formas diferentes, otro de “b” formas diferentes, y así sucesivamente, hasta que un acontecimiento puede ocurrir de “m” formas diferentes; entonces el total de formas que puede ocurrir un acontecimiento completo (tomando un solo elemento de cada conjunto), es dado por:

$$(a)(b)...(m)$$

- Principio de la suma

Un acontecimiento puede ocurrir de “a” maneras diferentes, otro de “b” maneras diferentes, y así sucesivamente, hasta que un acontecimiento puede ocurrir de “m” maneras diferentes. Suponiendo que los acontecimientos son mutuamente excluyentes(es decir si ocurre uno de los acontecimientos ya no pueden ocurrir otros); entonces cualquiera de los acontecimientos pueden ocurrir por la suma de las maneras; es decir,

$$a + b + ... + m$$

## 9.3. Ejercicios resueltos

1. Un estudiante para ir a la Universidad nacional de ingeniería (UNI), dispone de 8 líneas diferentes del metropolitano o de 4 líneas diferentes de microbús, de cuantas maneras diferentes se puede llegar a la UNI.

   Como los dos eventos son excluyentes(es decir si utilizo cualquier línea del metropolitano ya no podre usar cualquier línea de microbús y viceversa) el total de maneras es dado por la suma de cada acontecimiento:

   8+4=12

2. Sean los dígitos $\{1, \ 2, \ 3, \ 4\}$, se eligen 3 dígitos sin reemplazo, cuántos números de tres cifras se pueden formar.

Como el número a formarse es de tres dígitos, el primer digito se puede elegir de 4 maneras diferentes; como es sin remplazo el segundo digito se puede elegir de 3 maneras diferentes y por último el tercer digito se puede elegir de 2 maneras diferentes. Por lo tanto el total de números de tres dígitos que se pueden formar es dado por $(4)(3)(2) = 24$ (principio de la multiplicación)

Por formula $V_3^4 = \frac{4!}{(4-3)!} = \frac{4x3x2x1}{1} = 24$

Si la elección es con reemplazo, por el principio de la multiplicación (4)(4)(4)=64

Por formula $n^r = 4^3 = 64$

3. Sean los dígitos $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, se eligen 3 dígitos al azar y con reemplazo para formar un número. Hallar el total de números que se pueden formar en los siguientes casos:

a. Formar todos los posibles números.
Como las extracciones son con reemplazo, el primer, segundo y tercer digito pueden ocurrir de 9 formas diferentes, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 9 | 9 | 9 |

El total de formas diferentes de formar números con tres dígitos es dado por la multiplicación de 9x9x9=729.

b. Formar números pares
Un numero es par cuando la ultima cifra es par o cero, entonces el último digito puede ocurrir de 4 formas diferentes ( 2,4,6,8) y como es con restitución el primer y segundo digito pueden ocurrir de 9 formas diferentes, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 9 | 9 | 4 |
| | | |

El total de número pares que se pueden formar con tres dígitos es dado por la multiplicación de 9x9x4=324.

c. Formar números mayores que 500
El primer digito puede ocurrir de 5 formas diferentes para que sea mayor a 500 (5,6,7,8,9), y como es con restitución el segundo y tercer digito pueden ocurrir de 9 formas diferentes, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 5 | 9 | 9 |

El total de números que se pueden forman con tres dígitos mayores a 500 es dado por la multiplicación de 5x9x9=405.

d. Formar números de tal manera que el digito 3 se encuentre en el centro.
El primer digito puede ocurrir de 9 formas diferentes, el segundo digito puede solo ocurrir de una 1 sola forma (corresponde al número 3) y tercer puede ocurrir de 9 formas, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 9 | 1 | 9 |

El total de números que se pueden formar, de tal manera que el número 3 se encuentre en el centro es dado por la multiplicación de 9x1x9=81.

4. Sean los dígitos $\{1,2,3,4,5,6,7,8,9\}$, se eligen 3 dígitos al azar y sin reemplazo para formar un número. Hallar el total de números que se pueden formar en los siguientes casos:

a. Formar todos los posibles números
Como las extracciones son sin reemplazo, el primer, segundo y tercer dígitos pueden ocurrir respectivamente de 9, 8 y 7 formas diferentes, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 9 | 8 | 7 |

El total de números con tres dígitos es dado por la multiplicación de 9x8x7=504.

b. Formar números pares
Un numero es par cuando la ultima cifra es par o cero, entonces el último digito puede ocurrir de 4 formas diferentes

( 2,4,6,8) y como es sin restitución el primer y segundo digito pueden ocurrir de 8 y 7 formas diferentes, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 8 | 7 | 4 |

El total de número pares que se pueden formar con tres dígitos es dado por la multiplicación de 8x7x4=224.

c. Formar números mayores que 500
El primer digito puede ocurrir de 5 formas diferentes para que sea mayor a 500 (5,6,7,8,9), y como es sin restitución el segundo y tercer digito pueden ocurrir de 8 y 7 formas diferentes, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 5 | 8 | 7 |

El total de números que se pueden forman con tres dígitos mayores a 500 es dado por la multiplicación de 5x8x7=280.

d. Formar números de tal manera que el dígito 3 se encuentre en el centro
El primer digito puede ocurrir de 8 formas diferentes, el segundo digito puede solo ocurrir de una 1 sola forma (corresponde al digito 3) y el tercer puede ocurrir de 7 formas por ser sin restitución, entonces por el principio de la multiplicación se tiene:

| | | |
|---|---|---|
| 8 | 1 | 7 |

El total de números que se pueden formar, de tal manera que el número 3 se encuentre en el centro es dado por la multiplicación de 8x1x7=56.

5. En una reunión asisten 4 varones y 6 mujeres se elige 4 personas sin tomar en cuenta el orden y sin restitución, cuantos grupos de personas se pueden formar en los siguientes casos:

a. Todos los posibles grupos que se pueden formar.

$$C_4^{10} = \frac{10!}{4!*6!} = 210$$

b. Todos los posibles grupos en la cual existan 3 varones y 1 mujer.

$$C_3^4 * C_1^6 = 24$$

c. Todos los posibles grupos en la cual existan 2 varones y 2 mujeres.

$$C_2^4 * C_2^6 = 90$$

d. Todos los posibles grupos en la cual dos personas no aparezcan juntas

$$C_4^{10} - 1 * C_2^8 = 210 - 28 = 182$$

6. En una urna contiene 2 bolitas rojas, 1 verde y 1 negra, se eligen 4 bolitas sin reemplazo. Determinar el total de grupo a formarse

$$\frac{4!}{(2!)(1!)(1!)} = 12$$

9.4. Ejercicios propuestos

1. Sean los dígitos $\{0,1,2,3,4,5,6,7,8,9\}$, se eligen 4 dígitos al azar y con reemplazo para formar un número. Hallar:
   a. ¿Cuántos números pares se pueden formar.
   b. ¿Cuántos números son mayores a 5000.
   c. ¿Cuántos números son menores que 7000.
   d. ¿Cuántos números son mayores a 3000 pero menor a 8000.
   e. ¿Cuántos números son mayores a 3500.
   f. ¿Cuántos números contienen al digito 5.

2. En un estante existe 3 libros de historia, 4 de literatura y 2 de música, se eligen 2 libros sin reemplazo. Hallar todas las posibles combinaciones.

3. En una reunión asisten 5 varones y 3 damas, se elige 2 personas al azar. Hallar todas las posibles combinaciones.
4.
5. Un bibliotecólogo para codificar revistas científicas dispone de los siguientes dígitos $\{0,1,2,3,4,5,6,7,8,9\}$, de cuantos códigos de 4 dígitos se dispone en cada caso:

   a. Se toma en cuenta el orden y con reemplazo
   b. Se toma en cuenta el orden y sin reemplazo
   c. No se toma en cuenta el orden y sin reemplazo

6. Un cliente del Banco de la Nación tiene la posibilidad de escoger 4 de los siguientes 10 dígitos { 0,1,2,3,4,5,6,7,8,9} como clave de su tarjeta Multired, Hallar todas las posibles combinaciones bajos las siguientes restricciones:

   a. Esta permitido elegir un mismo digito( elección con reemplazo)
   b. No esta permitido elegir un mismo digito( elección sin reemplazo)
   c. Que el número elegido sea par( con reemplazo y sin reemplazo)
   d. Que el número elegido sea impar( con reemplazo y sin reemplazo)
   e. Que el número elegido sea superior a 5000( con reemplazo y sin reemplazo)
   f. Que el número elegido sea inferior a 5000( con reemplazo y sin reemplazo)
   g. Que el número elegido se encuentre entre 5000 y 8000( con reemplazo y sin reemplazo)
   h. Que el primer digito sea 4( con reemplazo y sin reemplazo)
   i. Que el último digito sea 4( con reemplazo y sin reemplazo)
   j. Que el segundo digito sea 0( con reemplazo y sin reemplazo)

7. En una reunión de una prestigiosa empresa dedicada a la producción de llantas para automóviles asisten 12 varones y 8 mujeres, después de acaloradas discusiones se desea formar una comisión para analizar la estrategias de ventas: Dicha comisión debe estar integrada por 5 personas. Hallar todas las

posibilidades de elegir a las 5 personas bajo las siguientes condiciones.

a. Elegir a cualquiera de ellos
b. Que todos sean del sexo varón
c. Que todos sean del sexo mujer
d. Que haya 3 varones y 2 mujeres
e. Que una mujer presida el grupo
f. Que un varón presida el grupo
g. Que una mujer este excluida del grupo
h. Que un varón este excluido del grupo

8. En una urna hay 4 bolas blancas, 5 rojas y 3 amarillas, se selecciona en forma aleatoria 5 bolas, cuantas formas existen de elegir las 5 bajo las siguientes modalidades:
   a. 2 sean blancas y 3 de otro color, sin reemplazo y sin considerar el orden de extracción.
   b. 3 sean blancas y 2 de otro color , con reemplazo y sin considerar el orden de extracción
   c. 2 sean amarillas y 3 de otro color, con reemplazo y considerando el orden de extracción.

9. Una persona para llegar a cierta ciudad tiene que hacer uso de los siguientes medios de transportes; tiene 3 líneas de ómnibus para llegar a determinado paradero y 4 líneas de ómnibus para llegar al paradero final. ¿De cuántas formas diferentes puede llegar al paradero final?.

10. En una urna existen 4 bolas rojas y 3 bolas blancas, se elige al azar 3 bolas con reemplazo. ¿Cuántas formas diferentes existen de elegir las 3 bolas?.

11. En un estante existen 5 libros de física de diferentes autores, 4 de matemáticas de diferentes autores y 3 libros de químicas de diferentes autores. ¿ cuantas formas existen de elegir tres libros de diferentes materias?.

12. De la pregunta anterior, Hallar las formas de elegir tres libros al azar con reemplazo.

13. De la pregunta 10, Hallar las formas de elegir tres libros al azar sin reemplazo.

14. De la pregunta 10, Hallar las formas de elegir tres libros al azar sin con reemplazo, si un libro de física siempre debe aparecer en el centro.

15. De la pregunta 10, Hallar las formas de elegir tres libros al azar con reemplazo, si un libro de física siempre debe aparecer en el centro.

16. En una biblioteca concurren 50 alumnos, de los cuales 30 son varones, se eligen al azar 10 alumnos, de cuantas formas se pueden elegir 5 varones y 5 damas.

17. Sean los dígitos $\{1, \quad 2, \quad 3, \quad 4\}$, se eligen 3 dígitos sin reemplazo, cuántos número de tres cifras menor o igual a 231 se pueden formar

18. En una biblioteca concurren 50 alumnos, de los cuales 30 son varones, se eligen al azar 10 alumnos, de cuantas formas se pueden elegir 6 varones y 4 damas.

# CAPÍTULO 10
# PROBABILIDAD

## INTRODUCCIÓN

En este capítulo se presentan las bases, para modelar situaciones donde interviene incertidumbre en la toma de decisión para un conjunto de datos observados.

## 10.1. EXPERIMENTO ALEATORIO $(\varepsilon)$

Es una operación física o idealizada, cuyo resultado de la observación no se conoce con exactitud, hasta después que culmine el experimento. El experimento deja de ser aleatorio cuando este concluye y se observa el resultado.

Ejemplos.

$\varepsilon_1$ =Elegir dos bolitas de una urna que contiene 1 bolitas rojas y 1 blancas.

$\varepsilon_2$ =Lanzar una moneda dos veces y observar su resultado

$\varepsilon_3$ =Registra el tiempo (horas) de vida de un foco eléctrico y observar su resultado.

$\varepsilon_4$ = Lanzar una moneda hasta que aparezca la primera cara.

$\varepsilon_5$ = Registra el peso (Kg.) de un recién nacido.

Características de un experimento aleatorio.

a. Puede ser repetido infinita veces, bajo las mismas condiciones
b. Los resultados posibles de un experimento, se pueden conocer a priori; Es decir se puede enumerar todos los posibles resultados del experimento aleatorio.
c. Para un experimento aleatorio es casi siempre posible establecer un modelo probabilístico.

## 10.2. ESPACIO MUESTRAL $(\Omega)$

Es el conjunto de todos los posibles resultados de un experimento aleatorio y es denotado por $(\Omega)$ *y* al total de resultados se denota por $n(\Omega)$

Ejemplos:

$\varepsilon 1 =$ Elegir dos bolitas de una urna que contiene 1 bolitas rojas y 1 bolitas blancas y anotar sus resultados.

$$\Omega_1 = \{RR, RB, BR, BB\}$$

$\varepsilon_2 =$ Lanzar una moneda dos veces y anotar sus resultados

$$\Omega_2 = \{(cc),(cs),(sc),(ss)\}$$

$\varepsilon_3$ = registra el tiempo (horas) de vida de un foco eléctrico.

$$\Omega_3 = \{t/t, \quad t \geq 0\}$$

$\varepsilon_4$ = lanzar una moneda hasta que aparezca la primera cara.

$$\Omega_4 = \{c, sc, ssc, sssc, ssssc, ...\}$$

$\varepsilon_5$ = Elegir a un recién nacido y registrar su peso (Kg.)

$$\Omega_5 = \{x/x = peso, \quad 1.5 \leq x \leq 4.1\}$$

## 10.3. TIPOS DE ESPACIO MUESTRAL:

ESPACIO MUESTRAL DISCRETO

Si el espacio muestral tiene un número finito o infinito numerable de elementos.

Ejemplo:

Son los espacios muéstrales $\Omega_1$, y $\Omega_2$

ESPACIO MUESTRAL CONTINUO

Si el espacio muéstral tiene un número infinito no numerable de elementos.

Ejemplos:

Son los espacios muéstrales $\Omega_3 , \Omega_4 , \Omega_5$

## 10.4. EVENTOS O SUCESOS

Es un subconjunto del espacio muéstral, o es el conjunto de todos posibles resultados que le interesa en forma particular al investigador. A los eventos se le representa por lo general con las primeras letras mayúsculas del abecedario o también con letra mayúscula acompañada con subíndice.

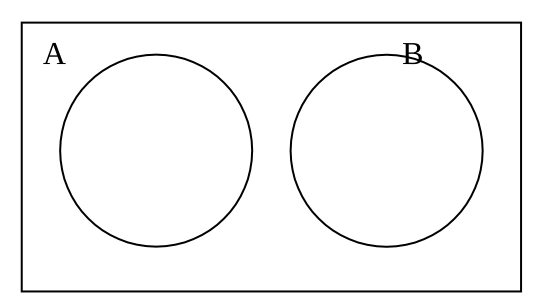
Ejemplo:

De los 5 espacios muéstrales anteriores se definen los siguientes eventos:

A = Que el número de artículos sea mayor a 5.
B = Obtener una cara en los dos lanzamiento.
C = Que el tiempo de vida se encuentre entre 980 horas y 1200 horas
D = Que ocurra cara en el 5to. Lanzamiento.
E = Que el peso del recién nacido se encuentre entre 2.5 y 3.4 Kg.

## 10.5. EVENTOS MUTUAMENTE EXCLUYENTES

Sean A y B dos eventos definidos en el espacio muestral $\Omega$ , se dice que los eventos son mutuamente excluyentes, si y solo si, estos no pueden ocurrir simultáneamente; es decir si el evento "A" ocurre, y no ocurre el evento B (viceversa); se representa matemáticamente por: $A \cap B = \Phi$



Ejemplo

Lanzar un dado dos veces, sean los eventos
A = Que la suma de sus puntos sean 7
B = Que la suma de su puntos sean 11

Luego los eventos A y B son mutuamente excluyentes; es decir si ocurre el evento A ya no ocurre el evento B, y viceversa.

TIPOS DE EVENTOS

10.6. EVENTOS SIMPLES

Son aquellos eventos que tienen un solo elemento del espacio muestral

Ejemplo

$\varepsilon_3$ = Lanzar un dado dos veces.

$$\Omega = \begin{Bmatrix} 1,1 & 2,1 & 3,1 & 4,1 & 5,1 & 6,1 \\ 1,2 & 2,2 & 3,2 & 4,2 & 5,2 & 6,2 \\ 1,3 & 2,3 & 3,3 & 4,3 & 5,3 & 6,3 \\ 1,4 & 2,4 & 3,4 & 4,4 & 5,4 & 6,4 \\ 1,5 & 2,5 & 3,5 & 4,5 & 5,5 & 6,5 \\ 1,6 & 2,6 & 3,6 & 4,6 & 5,6 & 6,6 \end{Bmatrix}$$

Sea el evento simple:

A = Que la suma de los resultados sea 2= $\{(1,1)\}$

10.7. EVENTOS COMPUESTOS

Son aquellos, eventos que tienen mas de un elemento del espacio muestral

Ejemplo

Del ejemplo anterior del lanzamiento de dos dados
A = Que la suma sea 7= $\{(3,4),(4,3),(2,5),(5,2)(6,1),(1,6)\}$

10.8. EVENTOS NULOS O VACIOS

Son aquellos eventos que no tiene elementos del espacio muestral

Del ejemplo anterior del lanzamiento de dos dados

A = Que la suma sea 13= $\{\quad\} = \phi$

10.9. EVENTOS IGUALMENTE PROBABLES:

Dos o más eventos simples son igualmente probables, si tienen la misma probabilidad de ocurrir.

Ejemplo-1

Se lanza un dado dos veces, los 36 eventos simples de la forma $(x, y)$ son igualmente probables; es decir cada evento simple tiene la misma oportunidad de ocurrir.

Ejemplo-2

Una urna contiene 8 bolas rojas y 2 bolas blancas, se elige una bola al azar, los eventos simples serán cada uno de los resultados posibles en este caso roja o blanca, notamos que la probabilidad de sacar bola roja es más probable que sacar bola blanca, entonces los eventos simples no son igualmente probables.

10.10. DEFINICIÓN CLÁSICA DE PROBABILIDAD

Es la manera más antigua de calcular probabilidades de eventos igualmente probables.

Sea $\varepsilon$ , un experimento aleatorio, y $\Omega$ su espacio muestral, con $N(\Omega)$ resultados igualmente probables. Sea A un eventos definido en $S$ , con $N(A)$ Resultados igualmente probables; la probabilidad de que ocurra el evento A denotado universalmente por P(A), es un número positivo definido por:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{N\,°de\ elementos del\ evento\ A}{N\,°de\ elementos del\ espaciomuestral\ S}$$

Ejemplo 1

Del ejemplo de lanzar un dado dos veces, cual es la probabilidad de que la suma de sus resultados sea siete.

Sea el evento A = La suma de los resultados de lanzar un dado dos veces sea siete.

A = $\{(3,4),(4,3),(2,5),(5,2)(6,1),(1,6)\}$ entonces n(A)=6

Como N( $S$ )=36, corresponde a todos los resultados posibles al lanzar un dado dos veces.

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{N° de\ elementos del\ evento\ A}{N° de\ elementos del\ espacio muestral\ S} = \frac{6}{36} = \frac{1}{6} = 0.1667$$

Interpretación:

La Probabilidad de obtener la suma de los resultados 7 cuando se lanza un dado dos veces es 0.1667 o el 16.67%.

Ejemplo 2

A una reunión asisten 30 varones y 40 mujeres, se eligen a 10 personas al azar sin reemplazo. Hallar las siguientes probabilidades:

a) Elegir 6 varones y 4 mujeres
b) Elegir por lo menos 4 varones

Solución para (a)

$S =$ Todas las posibles formas de elegir 10 personas de un total de 70, sin reemplazo:

$$N(\Omega) = C_{10}^{70}$$

Sea el evento A de interés, todas las formas posibles de elegir 6 varones y 4 mujeres de un total de 70 personas, es dado por:

$$N(A) = C_6^{30} C_4^{40}$$

Luego P(A)= $\frac{n(A)}{n(\Omega)} = \frac{C_6^{30} C_4^{40}}{C_{10}^{70}} = 0.1368$

Solución para (b)

Sea el evento B elegir por lo menos 4 varones, significa que en el grupo de 10 personas elegidas, deben haber 4 varones y 6 mujeres; 5 varones y 5 mujeres; 6 varones y 4 mujeres; 7 varones y 3 mujeres; 8 varones y 2 mujeres; 9 varones y 1 mujeres o 10 varones.

$$N(B) = C_4^{30} C_6^{40} + C_5^{30} C_5^{40} + C_6^{30} C_4^{40} + C_7^{30} C_3^{40} + C_8^{30} C_2^{40} + C_9^{30} C_1^{40} + C_{10}^{30}$$

$$P(B) = \frac{n(B)}{n(\Omega)} = \frac{C_4^{30} C_6^{40} + C_5^{30} C_5^{40} + \dots + C_{10}^{70}}{C_{10}^{70}} = 0.702052255$$

## 10.11. DEFINICIÓN FRECUENCIA RELATIVA DE PROBABILIDAD

Si un experimento aleatorio se repite n veces bajo las mismas condiciones y si $n_A$ de esas veces que se repite el experimento ocurre el evento de interés A. Luego la frecuencia relativa del evento A es:

$$h_A = \frac{n_A}{n}$$

En el límite de las veces que se repite el experimento, la probabilidad de ocurrencia del evento A es:

$$\lim_{n\to\infty} h_A = \lim_{n\to\infty} \frac{n_A}{n} = P(A)$$

La probabilidad, es el valor en el cual se estabiliza la frecuencia relativa del evento después de haber repetido un número grande de veces.

Ejemplo

Suponiendo que en una garita de peaje transitan casi siempre durante el día 8,000 vehículos, de los cuales 4,000 son autos, 2,500 son camiones y 1,500 corresponde a ómnibus.

| Tipos de vehículos | Frecuencias | Frecuencia relativa |
|---|---|---|
| Autos | 4000 | 0.5000 |
| Camiones | 2500 | 0.3125 |
| Ómnibus | 1500 | 0.1875 |
| Total | 8000 | 1. 000 |

a. ¿Cuál es la probabilidad de que pase por la garita de peaje un camión?.

Sea el evento A = que pase un camión por la garita de peaje, luego:

$$P(A) = \frac{2500}{8000} = 0.3125$$

b. ¿Cuál es la probabilidad de que pase por la garita de peaje un camión o ómnibus?.

Sea el evento B = que pase un camión o ómnibus por la garita de peaje,

$$P(A) = \frac{2500+1500}{8000} = 0.5$$

10.12. DEFINICIÓN AXIOMATICA DE PROBABILIDAD

Sea $\varepsilon$ , un experimento aleatorio y $\Omega$ su espacio muestral asociado al experimento; sea A un evento de interés definido en el espacio muestral, la probabilidad de que ocurra el evento A, es un número real positivo denotado por P(A), que cumple los siguientes axiomas:

a. $0 \leq P(A) \leq 1$

b. $P(\Omega) = 1$

c. *Sean* $A_1$ $y$ $A_2$ eventos mutuamente excluyente definidos en $S$ , entonces se cumple que*:*

$$P(A_1 o A_2) = P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

El axioma (c) se generaliza para más de dos eventos mutuamente excluyentes, siempre y cuando que sean excluyentes dos a dos.

Los tres axiomas son las bases para los siguientes teoremas:

1. Si $\phi$, es el conjunto vacío, entonces la $P(\phi) = 0$

   Demostración

   Sea A un evento definido en el espacio muestral $\Omega$ , entonces se cumple que: $A \cup \phi = A$, pero A y $\phi$, son eventos mutuamente excluyentes; es decir $A \cap \phi = \phi$ , haciendo uso del axioma (c), se tiene que:

$$P(A \cup \phi) = P(A) + P(\phi) = P(A) \rightarrow P(\phi) = 0$$

2. Sea $A^c$ un evento complemento del evento $A$ , entonces se cumple que:

$$P(A^c) = 1 - P(A)$$

3. Sean A y B dos eventos definidos en el espacio muestral $S$ , supongamos que $A \subset B$ , entonces se cumple que:

$$P(A) \leq P(B)$$

4. Sean A y B dos eventos cualesquiera definidos en el espacio muestral $\Omega$ , entonces se cumple que, la probabilidad de que ocurra el evento A o el evento B o ambos a la vez A y B es dado por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Ejemplo

Sean los dígitos: $\{1,2, 3, 4, 5, 6, 7, 8, 9\}$; se elige un dígito al azar. Calcular las siguientes probabilidades:

a. Que el número elegido sea menor a 4 o superior a 7.
b. Que el número elegido sea par o superior a 6.
c. Que el número elegido no sea 4

Solución para (a)

Sean los eventos A=que el numero elegido sea menor a 4 ; B= que el numero elegido sea superior a 7, luego:

$$A = \{1,2, 3\}$$
$$B = \{8,9,\}$$

Como los eventos A y B son mutuamente excluyentes, luego:

$$P(A \cup B) = P(A) + P(B) = \frac{3}{9} + \frac{2}{9} = \frac{5}{9}$$

Solución para (b)

Sean los eventos A=que el número elegido sea par; B=que el número elegido sea superior a 6, luego:

$$A = \{2, 4, 6, 8,\}$$
$$B = \{7, 8 ,9\}$$

Como los eventos A y B no son mutuamente excluyentes, luego:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

La intercepción de A y B es

$$A \cap B = \{8\}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{4}{9} + \frac{3}{9} - \frac{1}{9} = \frac{6}{9} = \frac{2}{3}$$

Solución para (c)

Sean los eventos A=que el número elegido sea 4 y su complemento $A^c$ = que el número elegido no sea 4:

$$A = \{4\}$$

$$P(A^c) = 1 - P(A) = 1 - \frac{1}{9} = \frac{9-1}{9} = \frac{8}{9}$$

## 10.13. PROBABILILIDAD CONDICIONAL

Sea A y B dos eventos definidos en el espacio muestral $\Omega$. Supongamos que el evento B ya ocurrió. ¿Cuál es la probabilidad de que ocurra el evento A?. El valor de la probabilidad pedida se representa universalmente por P(A / B), y es definida por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad para \quad P(B) > 0$$

La $P(A/B)$, cumple con los 3 axioma de probabilidad.

Ejemplo:

Un centro de información tiene 11,590 revistas científicas, las cuales están clasificadas según la materia y su idioma.

| Materias / Idiomas | Matemáticas | Física | Química | Total |
|---|---|---|---|---|
| Inglés | 1200 | 900 | 850 | 2950 |
| Portugués | 700 | 1500 | 900 | 3100 |
| Español | 1000 | 980 | 1200 | 3180 |
| Francés | 800 | 970 | 590 | 2360 |
| Total | 3700 | 4350 | 3540 | 11590 |

Se elige una revista al azar y supongamos que la revista esta escrita en inglés, ¿cuál es la probabilidad de que sea de física?:
Sean los eventos:
A= revistas de física
B= revistas escritas en inglés

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

De la tabla:

La probabilidad de que sea de física y escrito en inglés:

$$P(A \cap B) = \frac{900}{11590}$$

La probabilidad de que este escrito en inglés:

$$P(B) = \frac{2950}{11590}$$

Luego:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{900}{2950} = 0.30508$$

Otra forma de obtener este resultado:

Como el evento B ya ocurrió, el espacio muestral original conformado por la 11,590 formas de elegir una revista con esas dos características se reduce a solo 2,950 revista escritas en inglés, de las cuales 900 están en ingles, entonces:

$$P(A/B) = \frac{900}{2950} = 0.30508$$

## 10.14. REGLA DE LA MULTIPLICACIÓN DE PROBABILIDADES

De la probabilidad condicional

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \rightarrow P(A \cap B) = P(B)P(A/B) \quad (1)$$

También se tiene que la probabilidad de que ocurra el evento B suponiendo que ya ocurrió el evento A, es dado por:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \quad \rightarrow P(B \cap A) = P(A)P(B/A) \quad (2)$$

De 1 y 2 se tiene que:

$$P(A \cap B) = P(B)P(A/B) = P(A)P(B/A)$$

Ejemplo

Del ejemplo anterior, se eligen al azar y sin reemplazo dos revistas, calcular las siguientes probabilidades:

1. Que la primera revista sea de física y la segunda de física ( equivale decir también a que ambas sean de física)

   Sean los eventos:
   A=Elegir una revista de física
   B=Elegir una revista de física

$$P(A \cap B) = P(A)P(B/A) = \frac{4350}{11590} * \frac{4349}{11589} = 0.140847539$$

   También:

$$P(Ambas\, revistas\, sean\, de\ fisica) = \frac{C_2^{4350}}{C_2^{11590}} = 0.140847539$$

2. Que la primera revista sea de física y la segunda de matemática

   Sean los eventos:
   A=Elegir una revista de física
   B=Elegir una revista de matemática

$$P(A \cap B) = P(A)P(B/A) = \frac{4350}{11590} * \frac{3700}{11589} = 0.11983$$

3. Del ejemplo anterior supongamos que se eligen tres revistas al azar y sin reemplazo, cual es la probabilidad de que la primera revista este en idioma inglés, la segunda en idioma inglés y la tercera en idioma español.

   Sean los eventos:
   A=Elegir la primera revista en idioma ingles
   B=Elegir la segunda revista en idioma ingles
   C=Elegir la tercera revista en idioma español

$$P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B) = \frac{2950}{11590} * \frac{2949}{11589} * \frac{3180}{11588} = 0.01777$$

## 10.15. REGLA DE LA MULTIPLICACIÓN PARA MAS DE DOS EVENTOS CUALESQUIERA

Sea $A_1, A_2, ..., A_n$ eventos definidos en el espacio muestral $S$ , entonces la probabilidad de que ocurran todos estos eventos en ese orden es dato por:

$$P(A_1 \cap A_2 \cap, ..., \cap A_n) = P(A_1)P(A_2 / A_1)P(A_3 / A_1 \cap A_2), ..., P(A_n / A_1 \cap A_2, ... \cap A_{n-1})$$

Con la condición de que $P(A_1 \cap A_2, ... \cap A_{n-1}) > 0$

## 5. PARTICIÓN DEL ESPACIO MUESTRAL

Sea $A_1, A_2, ..., A_n$ eventos mutuamente excluyentes definidos en el espacio muestral $\Omega$, entonces $A_1, A_2, ..., A_n$, constituyen una partición, si cumplen las siguientes condiciones:

*1.* $A_i = \phi \quad \forall_i$

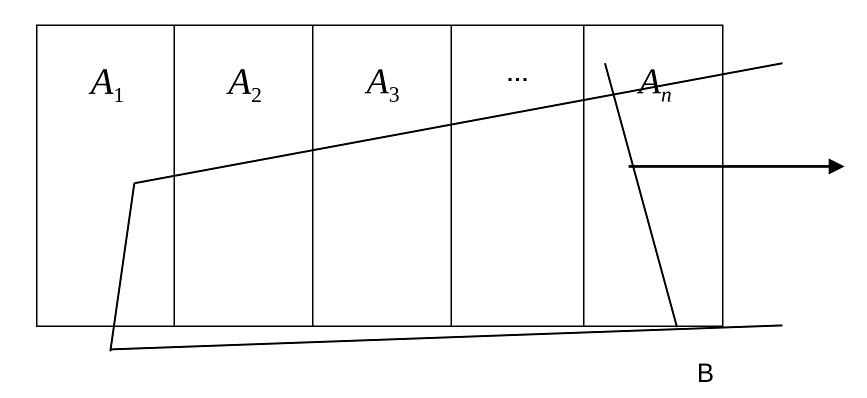
*2.* $A_i \cap A_j = \phi \quad \forall_{i \neq j}$

3. $A_1 \cup A_2 \cup ... \cup A_n) = \Omega$

| | | | | |
|---|---|---|---|---|
| $A_1$ | $A_2$ | $A_3$ | ... | $A_n$ |

## 10.16. TEOREMA DE LA PROBABILIDAD TOTAL

Sea $A_1, A_2, ..., A_n$ una partición del espacio muestral, definimos en estas particiones otro evento tal como B, luego se cumple que:




La probabilidad de que ocurra algún elemento de B es dado por:

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B)$$

Demostración:

Del GRÁFICO:

$B = (A_1 \cap B) \cup (A_2 \cap B) \cup ... \cup (A_n \cap B)$ , por ser estas intercepciones $(A_i \cap B)$ mutuamente excluyentes para $\forall_i$ luego se cumple que:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + ...P(B \cap A_n) = \sum_{i=1}^{n} P(A_i \cap B)$$

10.17. TEOREMA DE BAYES

Sea $A_1, A_2, ..., A_n$ una partición del espacio muéstral, definimos en esta partición otro evento B, entonces la probabilidad de que ocurra un evento $A_i$ suponiendo que ocurrió el evento B, es dado por:

$$P(A_i / B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B/A_i)}{P(B)} = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)} \quad i = 1,2,3,...n$$

Ejemplo

Un fabricante de chompas produce en un día determinado 1000 chompas; 300 son para varones, 250 para mujer y 450 para niños: La probabilidad de que una chompa de varón tenga falla en el tejido es de 0.01, para mujer es de 0.025 y para niños es de 0.05; de la producción de ese día se elige una chompa al azar; Hallar las siguientes probabilidades:

1. Que la chompa tenga falla
2. Que la chompa no tenga falla
3. Supongamos que se elige una chompa y tiene falla. ¿Cual es la probabilidad de que sea de varón?

De la información se tiene:

Sean los eventos:

$A_1$ =chompa de varón

$A_2$ =chompa de mujer

$A_3$ =chompa de niño

B = chompa con falla

Del enunciado se tiene las siguientes probabilidades:

$$P(A_1) = \frac{300}{1000} = 0.3$$

$$P(A_2) = \frac{250}{1000} = 0.25$$

$$P(A_3) = \frac{450}{1000} = 0.45$$

$$P(B/A_1) = 0.01$$
$$P(B/A_2) = 0.025$$
$$P(B/A_3) = 0.05$$

La probabilidad pedida es:

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3)$$
$$P(B) = 0.3 * 0.01 + 0.25 * 0.025 + 0.45 * 0.05 = 0.032$$

Interpretación:

Existe una probabilidad del 0.032 o 3.2% de elegir una chompa de la fábrica y que tenga falla.

La probabilidad pedida es:

$P(B^c) = 1 - P(B)$ =1-0.032=0.968 o 96.8%

La probabilidad pedida es:

$$P(A_1 / B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1)P(B/A_1)}{P(B)} = \frac{0.3 * 0.01}{0.032} = 0.094$$

## 10.18. INDEPENDENCIA DE EVENTOS

Sean A y B dos eventos definidos en el espacio muestral $\Omega$ , se dice que el evento A es independiente del evento B, si la probabilidad de ocurrencia del evento A, no se ve afectado por la ocurrencia del evento B y viceversa; es decir,

$$P(A/B) = P(A) \quad o \quad P(B/A) = P(B)$$

Por lo tanto, si los eventos A y B son independientes, entonces:

$$P(de\ que\ ocurra\ los\ eventos\ A\ y\ B\ en\ forma\ simultanea) = P(A\ y\ B) = P(A \cap B)$$
$$= P(A).P(B)$$

Teoremas

Sean A y B dos eventos independientes definidos en $\Omega$ , sean los eventos $A^c$ $y$ $B^c$ sus complementos respectivamente, luego se cumplen los siguientes teoremas:

1. $P(A^c \cap B) = P(A^c)P(B)$
2. $P(A \cap B^c) = P(A)P(B^c)$
3. $P(A^c \cap B^c) = P(A^c)P(B^c)$

Demostración de (3)

$A^c \cap B^c = (A \cup B)^c$ Por la Ley de Morgan

$$P(A^c \cap B^c) = P(A \cup B)^c = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B)$$

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A)P(B)$$

$$P(A^c \cap B^c) = (1 - P(A)) - P(B)(1 - P(A))$$

$$P(A^c \cap B^c) = (1 - P(A))(1 - P(B)) = P(A^c)P(B^c)$$

Ejemplo

Se tiene dos urnas la primera contiene dos bolas rojas y una bola blanca, la segunda urna contienen una bola roja y dos bolas blancas, se elige una bola de cada urna. ¿Cual es la probabilidad de ambas bolas sean del mismo color.

$$M = Obteneruna\ bola\ roja = \{R\}$$
$$N = Obteneruna\ bola\ blanca = \{B\}$$

Por ser ambos eventos independientes, luego:
Sea el evento de interés:

$$A = \{RR \quad o \quad BB\}$$

$$P(A) = P(\{RR \quad o \quad BB\}) = P(RR) + P(BB) = P(R).P(R) + P(B).P(B)$$

Por ser eventos mutuamente excluyentes y a la vez eventos independientes, luego se tiene:

$$P(A) = P(R).P(R) + P(B).P(B) = \frac{2}{3} * \frac{1}{3} + \frac{1}{3} * \frac{2}{3} = \frac{4}{9} = 0.44444$$

10.19. EJERCICIOS PROPUESTOS

1. Una caja contiene 8 bolsas de detergentes marca A y 5 bolsas de detergente marca B, se extraen al azar y con reemplazo 4 bolsas. ¿ Cuál es la probabilidad de que sean extraídas 3 bolsas del detergente A y 1 bolsa del detergente B?

2. En una reunión asisten 12 varones y 8 damas, se selecciona al azar 5 personas. ¿Cuál es la probabilidad de que 3 sean damas.

3. En una ciudad hay 3 vendedores de gas(A; B y C). Un día 5 residentes llaman a un vendedor de gas. Si cada residente selecciona al azar al vendedor de gas.

   a. ¿Cuáles la probabilidad de que todos los vendedores de gas sean elegidos.
   b. ¿Cuál es la probabilidad de que solo 2 vendedores de gas sean elegidos.

c. ¿Cuál es la probabilidad de que solo vendedores de gas A y B sean elegidos.

4. Se tienen los dígitos { 1,2,3,4,5,6,7,8,9} , se elige 3 dígitos al azar y con reemplazo. Hallar las siguientes probabilidades:

   a. Que el número elegido sea par
   b. Que el número elegido sea impar
   c. Que el número elegido sean mayor a 400
   d. Que el número elegido se encuentre entre 300 y 600
   e. Que el número elegido contenga al número 5 en el centro
   f. Que el número elegido contenga al número 5 al inicio
   g. Que el número elegido contenga al número 5 al final

5. Del ejemplo anterior Hallar las probabilidades para el caso de elegir los 3 dígitos al azar y sin reemplazo.
6. Si los dígitos {2,2,2,5,5,7,8}, se distribuyen al azar para formar un número de 7 cifras. ¿ Cuál es la probabilidad de que el numero formado se encuentren los dígitos 5 juntos?.
7. Una caja contiene 4 focos malos y 6 buenos, se sacan dos a la vez, si se prueba uno de ellos y se encuentra que es malo:¿Cuál es la probabilidad de que el otro también sea malo?
8. En un hotel se hospedan 6 arequipeños y 4 huanuqueño. Se elige en forma aleatoria a 3 personas del hotel. ¿Cuál es la probabilidad de que al menos un de ellos sea arequipeño?
9. El gerente de una tienda comercial de dedicada a la venta para varones, toma una muestra aleatoria de 200 clientes y lo organiza de acuerdo a la siguiente tabla de doble entrada, tal como se muestra:

| CLIENTES | COMPRAN A CRÉDITO | COMPRAN AL CONTADO | TOTAL |
|---|---|---|---|
| CON TARJETA DE CRÉDITOS | 80 | 40 | 120 |
| SIN TARJETA DE CRÉDITOS | 50 | 60 | 110 |
| TOTAL | 130 | 100 | 230 |

   a. ¿Cuál es probabilidad de que no tenga tarjeta de crédito o compre a crédito?
   b. ¿Cuál es la probabilidad de que tenga tarjeta de crédito y compre a crédito?

c. Si se sabe que tiene tarjeta de crédito. ¿Cuál es la probabilidad de que compre a crédito?
d. Se seleccionan a 3 clientes a azar, sin reemplazo. ¿Cuál es la probabilidad de que al menos uno compre al contado?
e. Se seleccionan a 3 clientes al azar con reemplazo.¿ Cuál es la probabilidad de que 2 tengan tarjetas de crédito?

10. Una compañía de cerveza tiene dos lavadoras de botellas. La lavadora “A” procesa un 20% de todas las botellas utilizadas diariamente y rompe un 3% de las que lava. La lavadora “B” procesa las restantes y rompe un 4%.

a. ¿Cuál es la probabilidad de que una botella lavada, seleccionada al azar este rota?
b. Una botella lavada seleccionada al azar, está rota. ¿Cuál a probabilidad de que haya sido lavada en la lavadora “A”?.
c. Una botella lavada seleccionada al azar, no está rota. ¿Cuál es la probabilidad de que haya sido lavada por la lavadora “A”?

11. Una base de datos contiene 50 libros de física y 20 de química y 30 de matemáticas, se eligen al azar 5 libros. Hallar las siguientes probabilidades.

a. Que 3 sean de Matemática y 2 de Química.
b. Que 2 sean de Química, 2 de Matemática y 1 de Física.
c. Que los 5 sean de Matemática.
d. Que 3 sean de Matemática y 2 de Física o los 5 sean de Matemática
e. Que los 5 libros sean de Física o 5 sean de Química.
f. Que los 5 sean de Matemática o 5 de Química o 5 de Física.

12. Sean los dígitos: 1, 2, 3, 4, 5, 6, 7, 8, 9, se eligen tres dígitos al azar con reemplazo para formar un número. Cuál es la probabilidad de que el número formado sea par o mayor a 500.

13. Una primera base de datos contiene 10 revistas de computación y 10 de física, otra segunda base contiene 8 de computación y 12 de física. Se elige una revista de cada base de datos. Calcular las siguientes probabilidades.

a. Que sean de la misma materia
b. Que sean de diferentes materias

# BIBLIOGRAFÍA


Baccini, Roberto Darío; Vásquez, Lara Viviana; Bianco, María José; Gracia Fonti; Javier (2018). Introducción a la probabilidad y a la estadística. Buenos Aires: Universidad de Buenos Aires.

Brase, Charles Henry (2019). Estadística básica. México: Cengage Learning.

Devore, Jay L. (2018). Fundamentos de probabilidad y estadística. México: Cencage Learning.

Hernández Sampieri, Roberto; Mendoza Torres, Christian (2018). Metodología de la investigación: las rutas cuantitativas, cualitativas y mixtas. México: McGraw Hill.

Mendenhall, William (2015). Introducción a la probabilidad y estadística. México: Cencage Learning.

Montanero Fernández, Jesús; Minuesa Abril, Carmen (2018). Estadística básica para las ciencias de la salud. Extremadura (España): Universidad de Extremadura.

Salazar P., Cecilia; Castillo G., Santiago del (2018). Fundamentos básicos de estadística. Recuperado de: http://www.dspace.uce.edu.ec/bitstream/25000/13720/3/Fundamentos%20B%C3%A1sicos%20de%20Estad%C3%ADstica-Libro.pdf

Suarez I., Mario O. (2018). Interaprendizaje de estadística básica. Recuperado de: http://repositorio.utn.edu.ec/bitstream/123456789/8696/1/Interaprendizaje%20de%20Estad%C3%ADstica%20B%C3%A1sica%20 2018.pdf

Quezada Lucio, Nel (2014). Estadística con SPSS 22. Lima: Macro.

